



Journée des doctorants

Résumés des communications PhDays 2017

Préface



Les axes de recherche du laboratoire **LMCS** s'articulent autour de **Méthodes** (d'Optimisation, de Modélisation, de Classification, ... etc.) pour la **Conception des Systèmes** (d'Information, Embarqués, Communicants, et de Reconnaissance).

Depuis sa création, les doctorants effectuant leurs thèses au sein du laboratoire sont un moteur de l'activité recherche dans ses différentes équipes. Ainsi, il est stratégique pour le laboratoire d'accompagner ses doctorants dans leur mission de formation par la recherche.

Dans cette perspective, la journée des doctorants est un rendez-vous annuel essentiel dans la vie du laboratoire. Cette journée permet de faire un point d'étape dans le parcours de la thèse. L'objectif est d'assurer un meilleur suivi des thèses au niveau du laboratoire et d'émettre un avis, sur les demandes de réinscription.

Ce document recense l'ensemble des thèses en cours au sein du laboratoire, et donne ainsi un aperçu assez sommaire de l'activité de recherche au sein du laboratoire.

Directrice du Laboratoire
Mme Karima Benatchba

Table des matières

Towards a Competency-based Framework for Crisis Management	1
Un cadre conceptuel pour la conception des systèmes d'information d'aide à la décision	3
Approches Intelligentes hybrides et parallèles pour la détection et le diagnostic de défaillances dans les machines asynchrones	5
Modèles pour la spécification des besoins des décideurs pour la conception des entrepôts de données.....	9
Une approche d'analyse de sentiment de texte en dialecte arabe basée sur la traduction automatique.....	10
Towards New Approach for the Enhancement of Adaptive Video Streaming	12
Un framework pour la fouille visuelle des données basé sur l'architecture Peer-to-Peer & la technologie des jeux vidéo online	14
Sécurisation des dispositifs biométriques mobiles.....	16
Un système de recherche d'information adaptatif dans un contexte d'intelligence économique	18
Simulation de la valeur des BMO (Business Model Ontologies) en entreprise par les SMA..	20
Framework d'analyse de Big Data basé sur la théorie des ensembles flous	22
Approches de combinaison d'informations biométriques de l'empreinte palmaire et digitale	24
Development of heterogeneous mapping approaches on Networks on chip.....	26
Indexation et recherche d'images par fusion d'informations multimodales	28
Impact des réseaux sociaux d'entreprise dans le partage des informations dans les organisations	30
Le routage dans les réseaux sur puce (NoCs).....	32
Un système de recommandation de contenu basé sur l'analyse du contenu social.....	34
La sécurité dans les réseaux centrés contenu	36
Social Computing for Enhancing Collective Intelligence	38
Optimal Deployment and Setting of Wireless Sensor Nodes for Energy Management in Smart Buildings	40
Integration spatio-temporelle de données et connaissances pour la prevention des risques professionnels en santé.....	42
Conception et génération d'un Meta-Modèle de connaissances à base de trace pour l'ingénierie des EIAH	44
Un système de recommandation des ressources Scientifiques.....	46
Approche intégrée pour l'ordonnancement de la production et la maintenance predictive basée sur le pronostic industriel	49
Publication de Données Massives Préservant l'Information Sensible	51

De la coopération des productions collectives dans un contexte multi-agent: un cas d'étude de surveillance de feux de forêts	53
Exploitation des réseaux sociaux pour l'amélioration des performances du BPM.....	55
Architectures flexibles et reconfigurables pour l'implémentation de la cryptographie à base des couplages	61
Cyber-Sécurité dans les Systèmes Cyber-Physiques	63
Vers un modèle des déterminants de l'usage efficace des <i>DSSs</i> : Application aux entrepôts de données.....	65
Modèles résiliables en biométrie par empreintes digitales.....	68
Capitalisation des connaissances d'entreprises dans un contexte d'économie du savoir par exploitation des technologies du e-Learning.....	69
Vers une Holonisation des services de e-gouvernement : étude des relations G to C et G to B	71
Approche d'ingénierie des connaissances en situation de collaboration intra et inter organisationnelle	73
Détection de communautés dynamiques par une approche d'analyse des réseaux sociaux	75
La théorie des jeux pour la résolution du problème d'affectation de fréquences dynamique dans un réseau GSM.....	77
Systèmes multi-agents ultra large échelle : contributions à l'introduction dynamique de systèmes dans un système de systèmes	80
Modèles de Représentation et d'Analyse des Réseaux Sociaux	83
Synergies entre optimisation combinatoire et data mining	85
Incorporation des données non-relationnelles dans les systèmes d'information décisionnel	88
La reconnaissance de l'empreinte palmaire latente.....	90
Systèmes de contraintes géométriques : décomposition et résolution	92
Energy efficiency in new generation networks	94
Considering learners emotions in Serious Games for eLearning context: a new approach	96
Etude de mouvement dans une séquence vidéo	98
Communications security in the next mobile networks generation	100
Analyse d'images satellites et photos aériennes multicritères pour la détection de changement sur des scènes urbaines : application à la mise à jour des bases de données géographiques .	102
Security in Big Data. Case Study: Privacy Protection	104
Adaptation du processus KDD (Knowledge Discovery in Databases) pour une utilisation dans un environnement Big Data.....	106
Prise de décision temps réel dans un contexte de Big Data	108
Collecte, Transport et Traitement (CTT) des données de masse dans un environnement mobile et incertain	110
Système Virtuel Sécurisé de gestion de clés de chiffrement dans le Cloud-Computing	114

Contributions à la détection et à l'agrégation des événements dans les microblogs.....	116
Approche de securisation des communications dans l'internet des objets (IoT)	118
Development of Multiobjective application mapping for heterogeneos network on chip	120
Une Ontologie de domaine pour l'aide à la décision dans un cadre juridique.....	122
Analyse des sentiments appliquee aux microblogs	124
Mesure de la satisfaction du décideur	126
Robustesse de la reconnaissance d'activité	128
Optimisation des techniques analytiques dans un contexte Big Data	131
L'Apprentissage des Agents dans un Environment Orienté Education	133
Assistance des apprenants dans un jeu sérieux Basée sur l'analyse de traces	135
Un accès contextuel adaptatif et proactif aux systèmes d'aide de décision mobile: approche basée sur l'utilisation d'un modèle flexible d'un profil contextuel sémantique.	137
Vers un Simulateur de Conception Physique de Bases de Données Sémantiques.....	139
Construction automatique des espaces de réponse dans les environnements multi-sources..	145
La Couverture dans les réseaux de capteurs par les théories de l'incertain	147
Location privacy in the Internet of things	149
Design of a system for the diagnosis of diseases based on the clustering of genomic data...	151
Une approche multidimensionnelle pour améliorer les systèmes de recommandation.....	153
Approches réalistes pour le déploiement et le contrôle des réseaux de capteurs sans fil.....	155
Approche générale pour la détection des cyber-attaques	161
Système crypto-biométrique pour la sécurité de l'information	163
Conception et développement de mécanismes à base des systèmes multi-agents tolérants aux fautes pour l'Adaptation dynamique des réseaux virtuels	165
Apport des informations imparfaites dans les systèmes de détection d'intrusions	167
Exploration et exploitation des aspects émotionnels chez l'apprenant en vue d'une définition d'un profil émotionnel pour un meilleur apprentissage en eLearning	170
Portail Sémantique pour la Gestion de Connaissances Destiné au Corpus (en langue arabe) des Hadiths du Prophète.	172
Internet of things Quality of Service Optimization Model	174
Big Graph Indexating and Compression	182
Vers Une plate-forme Open source pour le traitement automatique de la langue arabe.....	184
Approches biomimétiques parallèles et coopératives pour la résolution du problème de la coloration par listes. Application au problème d'affectation de fréquences.	187
Hyper-Heuristiques par selection et génération pour le probleme de Flowshop	189
Towards a reference ontology for information technologies	191
Annotation automatique d'images	193

Modèles et outils d’annotation pour une mémoire collective entre le décideur et le veilleur dans un contexte d’intelligence économique.	198
Infrastructure de communication sans fil avec qualité de service pour la gestion de crise et catastrophe.....	200
Critères de Performance et indicateurs pour évaluer le niveau d’apprentissage chez les apprenants dans un EIAH.....	202
Des méthodes biomimétiques pour les Problèmes d’Affectation de Fréquences (PAF)	204
Sécurité et économie d’énergie dans l’Internet of Everything	206
Resolution du problème d’affectation de fréquences multi objectifs dynamique en utilisant plusieurs approches biomimetiques	208
Modèles et Algorithmes Skyline pour la Sélection des Services Web Basés sur l’Historique d’Invocation	210
Analyse des réseaux sociaux pour la détection d’harcèlement	212
Ordonnancement conjoint de la production et de la maintenance avec contrainte d’affectation des ressources humaines dans les ateliers de production	214
Implémentation de techniques avancées d’Interaction Homme Robot pour un robot de service.....	216
Using background knowledge to enhance biomedical ontology matching.....	218
Un Modèle de Services Multifonctions pour les Systèmes Informatiques Ubiquitaires : Application au domaine Médical	220
Une approche d’aide à la constitution d’une source de données hétérogènes	222

Towards a Competency-based Framework for Crisis Management

Mohammed BENALI

Directeurs de thèse: Pr. Abdessamed Réda GHOMARI

Équipe: MSI

Abstract— Crisis events put the crisis response organizations in a unique and complex situation that requires critical real-time distributed decision-making so that lives and properties are saved and protected. Making efficient collective decisions needs an accurate awareness of the event context, and strongly depends on the effective use and coordination of resources, people, and information, where information and knowledge are obtained and owned by either organizations and even non-crisis experts people. With the growing development of collaborative technologies, citizens' participation to the crisis management process have shifted from the passive one-way contribution of social networking data to more active participation by performing specific tasks related to crisis-data processing. We intend by the mean of our research work to propose a comprehensive approach for integrating crowdsourcing processes and techniques to the collaborative decisional process in crisis situations. In this regard, this paper summarizes our research questions, the achieved work, and our future tendencies.

I. INTRODUCTION

In time of crisis, various response teams from different organizations need to collaborate and coordinate their efforts in attempt to conduct effective response operations for saving lives and properties. A process by which, crisis response organizations (CROs) seek to obtain a mutual awareness of the relevant crisis contextual conditions, collect and exchange real-time critical information, and share response action plans and resources for purpose of making efficient collective decisions. Crisis management implies collaborative, distributed and complex decision-making activities involving diverse sources of spatiotemporal information and knowledge, distributed among stakeholders from different organizations, agencies and even ordinary citizens. With the emergence of the web 2.0 and the increasing development of collaborative technologies, citizens' engagement has shifted from the simple providing of raw data and information such as social media posts, towards more established contribution by performing specific tasks related to information management. This emerging participatory tasking process is known as "Crowdsourcing phenomenon". In this regard, [1] underlines the power of recent advances in technologies and their capacity to harness the collective intelligence on greater scale than ever before. He claims that crowdsourcing, wisdom of the crowds' concepts, social networks, collaborative software, and other web-based tools constitute a shift paradigm in the way that organizations make decision and calls this emerging era: Decision 2.0.

II. PROBLEM STATEMENT

Recent crisis and disaster events have led respondent organizations to reconsider their policies so as to expand the range of collaborating actors, by including volunteer citizens as well as expert operators, to support their internal decision-making activities [2]. In unexpected problems, especially crisis situations, it has been noticed that decision authority flows down to those closest to the situation in either location or knowledge [3]. In this regard, there is still more research questions and studies of practical scenarios to be addressed by crisis informatics scholars and practitioners, in order to analyze and examine the engagement and the role that ordinary citizens may play in the different steps of decision-making activities in crisis situations. In this perspective, the main preoccupation of research is find new ways of understanding, conceptualizing and defining specific guidelines and design methodologies for integrating the relevant crisis crowdtasking activities and techniques with respondent organizations' decisional process that translates into resources allocation and prioritization of response and relief actions.

To the best of our knowledge, current literature and studies on the topic are lacking a concise approach for integrating crowdsourcing processes to organizations' decisional processes, and most of the authors address considerations and challenges related to aligning crowdsourcing with decision-making for general problem domain and did not examine in particular the crisis management field.

III. PERFORMED WORKS

A. First year

1) State of the art

a) *Crisis management: First, we have started by studying the crisis management domain, addressing its terminological ambiguities and outlining the fundamentals of the later (crisis classifications, crisis and disaster life cycles crisis management models, etc.).*

b) *Collaborative Crisis management aspects (crisis scales, crisis phases, crisis management support tools):*

c) *Collaborative Crisis Management Issues: in this point, we have examined the multitude of challenges and barriers that face the involved collaborating stakeholders, considering their significant discrepancies and differences in terms of cultures, skill sets, objectives and priorities.*

- *Interaction issues (communication, collaboration, situation and activity awareness).*

- *Improvisation* (ad hoc decision-making, cooperation).
- *Decision making issues.*
- *Collective Intelligence.*
- *Organizational learning* (how to leverage the lessons learned)

B. Second Year

1) First contribution (JUNE 2016)

Based on our bibliographic readings, we have proposed a meta modeling of the main literature thematic in the area of crisis and disaster management. This literature is then synthesized to develop a classification of information and knowledge driven collaborative crisis management works.

2) Published papers

In January 2017, we have submitted our second paper [5] to the 14th International Conference for Crisis Response and Management (**ISCRAM2017**). We have received the acceptance for presentation in March 2017. The paper has been indexed in the conference proceeding and the DBLP library.

C. Third year

1) Second contribution (JANUARY 2017)

The first contribution has allowed us to pin down the most significant research directions in the crisis management domain. More importantly, our bibliographic readings and analyses have revealed the lack of works which leverage the power of crowd (citizen's involvement) for decision support in crisis and disaster management. As a result, the second step of our research work was to examine how crowdsourcing can influence the different phases of the decision-making process in a collaborative crisis management situation. After discussing the literature revue, we have presented and discussed in our second contribution the proposal of a crowdsourcing-based approach for decision making that relies on Crowds participation for each phase of the decisional process.

2) Published papers

In June 2016, we have submitted our paper [4] as a short paper to the 3rd IEEE international conference on Information and Communication Technologies for Disaster Management (**ICT-DM2016**). We have received, in September 2016, the acceptance for presentation and inclusion of the paper in the conference proceedings.

IV. PERSPECTIVES

Within the second contribution, and for each step of Simon's decisional process (IDC), we specify the main crisis crowdsourcing tasks which the crowds are called to perform, and the appropriate technique for outsourcing those tasks. To exhibit the coordination work that exists between the response organizations and the involved citizens in the decision-making process, we use the Social extension of Business Process Model (BPMN4Social) that aims to define a specific notation for describing Social BPM behaviors. The main objectives of the BPMN4Social is making the internal decision procedures more visible to the affected stakeholders, assigning an activity to a broader set of contributors, and eliciting opinions that contribute to the decision-making process. However, developing a crisis management strategy may be considered as knowledge-intensive process driven by crisis knowledge

workers' own experience based on the emerging crisis circumstances and its unique features. Unfortunately, BPMN extended standards such BPMN4Social are typically used to model well-structured and automated routine processes. To cope with this issue, we adopt the Case Management Model and Notation (CMMN) standard [6] which has been developed to complement BPMN in modeling and managing of processes that depend on circumstances and ad-hoc decisions of knowledge workers.

In addition, another aspect that we intend to address in our proposal is the modeling of specific activities related to decision. In fact, modeling the decisional process using the BPMN4Social business process models for each phase can only describe the coordination work that exists between the crisis response organizations and the involved citizens within business processes by defining specific tasks or activities where decision-making is required to take place. However, these decision-related tasks or activities each have their own internal structure which is not well managed by traditional business process models, which is the case for BPMN4Social. Therefore, we adopt the Decision Model and Notation (DMN) standard [7] to model decision activities.

As a result, we intend for what follows to extend the achieved work (conference papers) so that to take into consideration the last two points discussed above. In addition, to demonstrate the applicability of our approach in a practical scenario, we are carrying out a study of the Desert Locust Plague in the Algerian National Institute for Plant Protection (NIPP).

V. CONCLUSION

Given the collaborative nature of crisis management, the idea behind crowdsourcing platforms is to allow a more established collaboration between different involved organizations and the large public. However, in reality, the platforms lack collaboration between the response organizations in term of planning and coordination of relief actions. In this regard, crowdsourcing applications must allow the possibility of integration of collaborative tools into the Crowdsourcing systems to overcome coordination barriers.

REFERENCES

- [1] E. Bonabeau, "Decisions 2.0: The power of collective intelligence". In: *MIT Sloan management review* 50.2, pp. 45, 2009.
- [2] A. Ginige, L. Paolino, M. Romano, M. Sebillio, G. Tortora, and G. Vitiello, "Information sharing among disaster responders-an interactive spreadsheet-based collaboration approach". In: *Computer Supported Cooperative Work* 23.4-6, pp. 547-583, 2014.
- [3] M. Turoff, C. White, L. Plotnick, and S. R. Hiltz, "Dynamic emergency response management for large scale decision making in extreme events". In: *Proceedings of the 5th International Conference on Information System for Crisis Response and Management. ISCRAM2008. Washington, DC, USA*, pp. 462-470, 2008.
- [4] M. Benali and A. R. Ghomari, "Information and knowledge driven collaborative crisis management: A literature review," in *3rd International conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. IEEE, 2016.
- [5] M. Benali and A. R. Ghomari, "Towards a Crowdsourcing-based Approach to enhance Decision Making in Collaborative Crisis Management" In: *Proceedings of the 14th International Conference on Information System for Crisis Response and Management. ISCRAM2017. Albi, France*, 2017.
- [6] OMG, "Case Management Model and Notation 1.1", 2016, <http://www.omg.org/spec/CMMN/1.1/PDF>.
- [7] OMG "Decision Model and Notation 1.1", 2016, <http://www.omg.org/spec/DMN/1.1/PDF>.

Un cadre conceptuel pour la conception des systèmes d'information d'aide à la décision

Mohammed El Amine Tali

Directrice de thèse : Mme Fahima Nader

Résumé— Ce document a pour but de présenter un résumé de l'état d'avancement des travaux de la thèse, qui sont axés sur la problématique de la conception des systèmes d'information d'aide à la décision. Les travaux réalisés, les résultats actuels ainsi que les travaux à venir seront détaillés dans les différentes sections de ce document.

I. INTRODUCTION

Les systèmes d'information d'aide à la décision sont largement définis comme des systèmes informatiques qui aident les décideurs à utiliser des données, des connaissances et des modèles pour résoudre des problèmes et prendre des décisions [1] [2]. Ces systèmes sont apparus pendant les années soixante, il y a eu depuis de nombreuses évolutions [3] [4]. Selon Sprague et Carlson [9] un SIAD comporte trois composants fondamentaux : une interface homme-machine, un Système de Gestion de Bases de Données (SGBD) et un Système de Gestion de Bases de Modèles (SGBM).

Nous considérons que chaque processus dans l'organisation nécessite un pilotage [5]. Et, qu'il doit être doté de systèmes d'information adaptés aux situations décisionnelles [6] auxquelles sont confrontées les pilotes de tels processus. Cependant la conception d'un système d'information d'aide à la décision reste toujours une tâche complexe et difficile qui se base sur l'expérience et les connaissances dans différents domaines (aide à la décision, statistique, recherche opérationnelle, analyse des systèmes d'information, IHM...) [7].

II. PROBLEMATIQUE

Les approches classiques du développement des systèmes d'information ne sont pas adéquates au processus de développement des SIAD, en raison des spécificités de ces derniers [8]. Quelques méthodes de conception de SIAD ont été proposées : ROMC [9], KADS (knowledge acquisition and documentation structuring) [10], prototypage, développement itérative,... Ces méthodes restent liées soit à des outils (entrepôts de données, ETL, cubes, requêteurs, ...), soit à un type particulier de SIAD (groupe, négociation, à base de connaissance, ...), soit à des domaines d'application (médical, finance, gestion des crises, environnement, ...), soit spécifiques à des projets. Aucune de ces méthodes n'est dominante, plusieurs auteurs [11] [12] [13] [14] affirment le manque d'une méthode générique pour la conception des SIAD.

III. TRAVAUX REALISES

A. Approche proposée

Nous travaillons sur la proposition d'une approche orientée-décision pour la conception des systèmes d'information d'aide à la décision. En effet, afin de pouvoir concevoir un SIAD,

nous devons d'abord étudier le système décisionnel pour bien comprendre la situation de décision réelle dans laquelle le décideur agit. Notre proposition part de l'hypothèse que développer un SIAD ne peut se faire sans s'intéresser aux problèmes de décision qui se posent aux différents utilisateurs finaux, au contenu des décisions, et au processus de prise des décisions. En effet les méthodes actuelles de conception de ce type de systèmes négligent l'aspect de décision lors de la conception et se focalisent dans la plupart du temps sur les données.

L'approche proposée se déroule en 4 étapes : l'identification des décisions, l'analyse des décisions, la conception et la maintenance. Quatre types de rôles sont impliqués dans la mise en œuvre l'approche : les décideurs, les analystes, les développeurs et les knowledge managers.

La première étape consiste à l'identification des décisions aux quel le décideur peut faire face au cours du pilotage d'un processus métier donné. Pour faire nous utilisons la grille GRAI [15], qui est un outil puissant pour identifier les points où les décisions sont prises. A la fin de cette étape on sera capable d'identifier tous les décisions, les informations internes et externes nécessaires pour prendre ces décisions ainsi que le niveau de ces décisions (Stratégique, tactique, opérationnel). Après avoir identifié toutes les décisions, l'étape suivante (l'analyse des décision) consiste à déterminer comment chaque décision est prise et quel sont les informations nécessaires pour prendre cette décision. Pour faire nous allons modéliser chaque décision [16] en utilisant le nouveau standard « Modèle de Décision et de Notation » DMN [17], proposé par l'OMG (Object Management Group). Le but de ce standard est de fournir une notation commune qui est facilement compréhensible pour tous les acteurs professionnels (analystes, développeurs, décideurs, ...). Le DMN fourni deux niveaux de modélisation de décision, le niveau d'exigences de décision pour définir la structure des décisions, les données exigées pour chaque décision et les relations entre les décisions, et le niveau logique de décision pour spécifier comment chaque décision est prise en fonction des données disponibles. A la fin de cette étape on pourra spécifier pour chaque décision les informations ainsi que le modèle nécessaire (règles métier, table de décision, algorithme, modèle analytique...) pour prendre cette décision. La troisième étape consiste à la conception de la base de données et la base de modèle, les deux composants principaux du SIAD.

Après avoir conçu la première version du SIAD, ce dernier doit être maintenu et mis à jour. En effet, cette première version prend en charge les décisions identifiées dans la première étape. Cependant, l'environnement décisionnel change et évolue rapidement. De nouvelles décisions à prendre ou d'informations ou de connaissances relatives à une décision déjà identifiée peuvent apparaître, et doivent être implémenter.

B. Production scientifique

- Une communication nationale : « Une grille de lecture des outils d'aide à la décision », 10ème Journées Scientifiques et techniques de Sonatrach, ORAN, Octobre 2015.
- Une communication internationale : « Decision Support Tools : A Novel Classification », In Proceedings of the 2nd EWG-DSS International Conference on Decision Support System Technology (ICDSSST 2016), Plymouth, United Kingdom, Mai 2016.
- Une communication international: « Decision modelling: Methods and tools ». Soumis à la conference « The International Conference on Mathematics and Information Technology ICMIT 2017, Adrar, Algérie».

IV. PERSPECTIVES

Notre objectif actuel est de compléter l'approche et rédiger un papier journal et le soumettre dans les plus brefs délais.

V. CONCLUSION

Diverses difficultés se posent toujours, lors de la conception des systèmes d'information d'aide à la décision. La proposition des méthodes pour développer ce type de système est un sujet qui a préoccupé la communauté des SIAD pendant longtemps. Quelques méthodes sont proposées mais aucune d'elles n'est dominantes. Nous tentons à palier a ce manque par la proposition d'une nouvelle approche orienté décision pour la conception des systèmes d'information d'aide à la décision basée sur la méthode GRAI et le modèle DMN. Notre travail n'est pas complètement terminé, nous visons l'extensibilité et l'enrichissement de notre approche.

References

- [1] D. J. Power, S. Ramesh, «Decision Support Systems», Handbook of Automation, 1539-1548, 2009.
- [2] G. A. Gorry, M. S. Morton, «A framework for management information systems», MIT Sloan Management Review, vol. 30, no 3, p. 49, 1989.
- [3] M. A. Tali, R. Chalal, «Une grille de lecture des outils d'aide à la décision», 10ème Journées Scientifiques et techniques de Sonatrach, ORAN, Octobre 2015.
- [4] M. A. Tali, F. Nader, R. Chalal, «Decision Support Tools: a Novel Classification», In Proceedings of the 2nd EWG-DSS International Conference on Decision Support System Technology (ICDSSST 2016), Plymouth, United Kingdom, Mai 2016.
- [5] J. Mélése, « L'analyse modulaire des systèmes (AMS), une méthode efficace pour appliquer la théorie des systèmes au management », Editions Hommes et Techniques, 1972.
- [6] R., Chalal, « Une approche pour la capitalisation coopérative des connaissances sur les risques produit en phase initiale d'un projet industriel », Thèse de doctorat, Institut National de Formation en Informatique, INI, Oued Semar- Alger, 2007.
- [7] D. Arnott, G. Pervan., «Eight key issues for the decision support systems discipline», Decision Support Systems, vol. 44, pp. 657–672,2008.
- [8] G.M. Marakas, «Decision Support Systems in the 21st century», 2nd edition, Prentice Hall, New Jersey, 2003.
- [9] J.R. Sprague, H. Ralph, E. D. Carlson «Building Effective Decision Support Systems», Prentice-Hall, Inc,Englewood Cliffs, 1982.
- [10] C. Vogel «Génie Cognitif», Masson, 1988.
- [11] M. Salles «Decision-Making and the Information System», Wiley, Vol. 3, Advances in information systems, juin 2015.
- [12] D. Fogli, G. Guida, «Knowledge-centered design of decision support systems for emergency management», Decision Support Systems, 2013, vol. 55, no 1, p. 336-347.
- [13] S. Lepreux, «Approche de Développement centré décideur et à l'aide de patrons de Systèmes Interactifs d'Aide à la Décision-Application à l'investissement dans le domaine ferroviaire», Thèse de doctorat. Université de Valenciennes et du Hainaut-Cambresis, 2005.
- [14] A. Gachet, R. Sprague, «A context-based approach to the development of decision support systems», Proceedings of the 5th International and Interdisciplinary Conference on Modeling and Using Context, Paris, France, 5–8 July 2005.
- [15] G. Doumeingts, B. Vallespir, D . Chen, «Decisional modelling GRAI grid», International handbook on information systems, Bernus P., Mertins K. and Schmidt G. ed., Berlin: Springer ,1998.
- [16] M. A. Tali, F. Nader, R. Chalal, «Decision modelling: Methods and tools», unpublished.
- [17] OMG, «Decision Model and Notation,» 2016.

Approches Intelligentes hybrides et parallèles pour la détection et le diagnostic de défaillances dans les machines asynchrones

KHAMOUDJ Charaf eddine

Directrice de thèse : Benbouzid-Si Tayeb Fatima (LMCS-ESI)
Co- Directeur de thèse: Benbouzid Mohamed El Hachemi (LBMS-UBO)

Résumé— Dans ce travail nous allons proposer une approche pour la détection et le diagnostic de défaillance dans les machines asynchrones au niveau de roulement. L'approche commence par une classification non supervisée basée sur les signaux de vibrations et de température mesurés par la plateforme d'expérimentation PRONOSTIA pour construire des classes qui représentent l'état de santé du système, puis d'appliquer une analyse des associations pour surveiller le fonctionnement du système.

I. INTRODUCTION

La détection et le diagnostic de défauts/défaillances dans les systèmes électromécaniques sont une priorité dans le cadre global de la sûreté de fonctionnement. Un des composants sensibles de ces systèmes est la machine électrique et en particulier la machine asynchrone qui est largement utilisé dans l'industrie en raison d'un grand nombre d'avantages ; entre autres sa simplicité de construction, son faible coût, sa résistance mécanique et son fonctionnement à vitesse variable. Cependant, son utilisation génère des contraintes électriques et mécaniques, sources de défaillances internes qui doivent être détectées puis diagnostiquées dès leur naissance. De plusieurs études statistiques, il en ressort que l'élément le plus sensible est le/les roulements [1]. Dans notre travail nous proposons une approche basée sur les signaux de vibrations et de température acquis de la plateforme PRONOSTIA [2-3], pour assurer le diagnostic de l'état de fonctionnement et la détection de défaillances des roulements. L'objectif de l'approche est d'arriver à une représentation graphique qui regroupe les signaux mesurés, où chaque groupe représente un état de fonctionnement pour assurer le diagnostic du système, puis, d'analyser les transactions d'un état à un autre pour former un modèle de transactions pour assurer la détection de défaillances. Pour la première phase, nous nous sommes basé sur la classification non supervisée de données pour construire les classes qui représentent les états de fonctionnement. Pour résoudre ce problème, nous allons hybrider les métaheuristiques VNS (Variable Neighborhood Search) [4] et CMO (Classical Mechanics Optimization) [5-6], où la deuxième est utilisée pour définir une solution initiale à la première. La deuxième phase consiste à créer un modèle graphique des classes obtenues dans la première phase, afin d'appliquer l'analyse des associations pour définir les transactions de l'état de roulement.

II. PROBLEMATIQUE

Le rôle essentiel d'un moteur électrique est la transformation de l'énergie électrique en énergie cinétique de

rotation. Il est composé en deux parties principales, le stator et le rotor:

- Le stator représente la partie fixe du moteur, c'est une carcasse en fonte sous forme d'une couronne, il contient plusieurs bobines qui jouent le rôle d'inducteur.
- Le rotor représente la partie rotative du moteur qui joue le rôle de l'induit. C'est un cylindre claveté sur le stator avec des entrefers. Il contient deux roulements placés sur les deux côtés.

Les défaillances les plus fréquentes des moteurs électriques sont au niveau des roulements. Il existe, dans la littérature, différentes méthodes de détection et de diagnostic. Ces méthodes sont basées sur deux types d'approche Les méthodes de la détection et le diagnostic de défaillance. Le premier groupe englobe les méthodes basées sur les modèles mathématiques pour décrire le fonctionnement normal du moteur électrique, puis, de prévoir la défaillance selon la perturbation du mode de fonctionnement par rapport au modèle mathématique. Le deuxième groupe englobe les méthodes basées sur les signaux mesurés pour la détection et le diagnostic de pannes, ces signaux permettent d'évaluer les conditions de fonctionnement du moteur asynchrone [7]. Dans ce dernier contexte, il est indispensable de surveiller l'état de santé du système qui nécessite un déploiement de capteurs et des techniques de traitement du signal. L'analyse des vibrations reste la technologie de surveillance la plus utilisée dans l'industrie [8]. Dans [9], une méthode d'évaluation pour la surveillance de la dégradation des performances des roulements basée sur le regroupement des k-medoids à l'aide des signaux de vibration collectés est proposée. Elle consiste à traiter les signaux en utilisant une mesure de similarité pour former une classification puis à obtenir les classes qui représentent les états de défaillance. [10] ont utilisé k-means dans la procédure d'apprentissage pour le diagnostic de défaillance des roulements. Vu que, k-means est sensible au choix des centres initiaux, [11] a proposé d'hybrider k-means avec un algorithme génétique pour définir la solution de départ pour implémenter une méthode de diagnostic de défaillance de roulement basée sur des signaux de vibration. [8] propose une description intéressante des algorithmes de détection et de diagnostic des pannes basés sur les signaux de vibration.

III. - TRAVAUX RÉALISÉS

A. Proposition d'une approche

Dans notre travail, nous allons proposer une nouvelle méthode pour la détection et le diagnostic de défaillance des

roulements basée sur les signaux mesurés de vibration et de température à partir de la plateforme PRONOSTIA. La solution proposée est composée en deux phases principales. La première phase consiste à grouper les signaux mesurés dans des classes en utilisant la classification non supervisée à l'aide de l'hybridation de deux métaheuristiques, CMO pour définir une solution de départ et VNS pour créer les classes. Après la construction des états, la deuxième phase consiste à appliquer une analyse des associations qui se base sur un algorithme rationnel qui définit les transactions des états de roulements d'un état à un autre. C'est-à-dire rechercher les modèles qui se produisent fréquemment dans les dégradations. L'extraction des modèles fréquents mène à la découverte d'associations intéressantes pour prévoir le prochain état du roulement en fonction des signaux mesurés, du temps et d'autres conditions comme la charge. Si le prochain état représente une défaillance donc nous le signalons avant son apparition.

B. Publication de la métaheuristique CMO

La métaheuristique CMO a été publiée dans «*Springer's Lecture Notes in Computer Science*» [5], comme post-proceedings, après la participation à la conférence ICSIBO 2016 (International Conference on Swarm Intelligence Based Optimization) [6] CMO a été utilisée dans la classification non supervisée pour résoudre le problème de la segmentation d'image. C'est une perspective du travail réalisé au magister qui consiste à l'application du CMO dans le Datamining.

C. Évaluation et validation de l'approche

La première phase de l'approche a été implémentée en utilisant CMO. Les résultats obtenus permettent de créer des classes qui représentent les états de fonctionnement (figure 1).

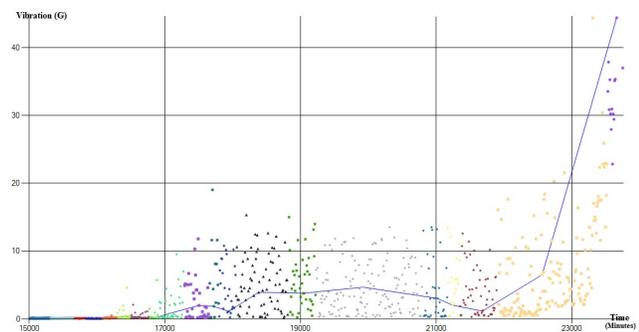


Fig. 1. Exemple d'une classification non supervisée des signaux de vibration.

Les résultats obtenus permettent aussi de suivre l'évolution de la dégradation des roulements dans le temps (figure 2).

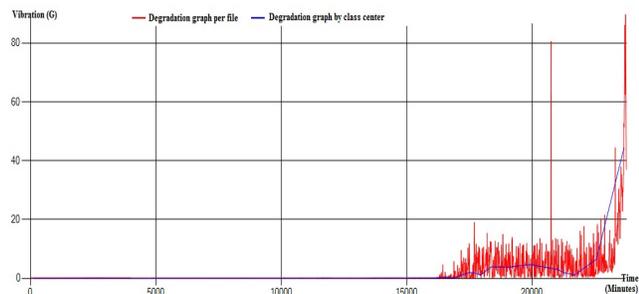


Fig. 2. Exemple du graphe de l'évolution de la dégradation du roulement.

IV. PUBLICATION DE L'APPROCHE DANS UNE CONFÉRENCE

Un article intitulé «*Classical Mechanics-Inspired Optimization Metaheuristic for Induction Machines Bearing Failures Detection and Diagnosis*» a été soumis à la conférence IECON 2017 «*43rd Annual Conference of the IEEE Industrial Electronics Society (IES)*». L'article a été accepté pour la participation à la conférence.

V. PERSPECTIVES

Les travaux à réaliser sont les suivants :

- Rédaction de l'état de l'art sur les techniques existantes.
- Implémenter et tester l'approche.
- Publication de l'approche globale dans une revue.
- Rédaction des parties finalisées.

Références

- [1] E. Elbouchikhi, V. Choqueuse and M.E.H. Benbouzid, "Current frequency spectral subtraction and its contribution to induction machines bearings condition monitoring," *IEEE Transactions on Energy Conversion*, vol. 28, n°1, pp. 135-144, Mars 2013.
- [2] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA: An experimental platform for bearings accelerated life test," in *Proceedings of the 2012 IEEE PHM*, Denver (USA), pp. 1-8, June 2012.
- [3] <http://www.femto-st.fr/en/Research-departments/AS2M/Research-groups/PHM/Pronostia>.
- [4] J. Brimberg, P. Hansen and N. Mladenovic, "Attraction probabilities in variable neighborhood search" Article in *4OR quarterly journal of the Belgian, French and Italian Operations Research Societies* 8(2):181-194, Juin 2010.
- [5] C. Khamoudj, K. Benatchba, and M. T. Kechadi, "Classical mechanics optimization for image segmentation," in *Proceedings of the 2016 ICSIBO*, Mulhouse (France), pp. 102-110, June 2016.
- [6] C. Khamoudj, K. Benatchba, and M. T. Kechadi, "Classical Mechanics Optimization for Image Segmentation," *ICSIBO 2016*, Springer LNCS, pp. 1-9, vol 10103, 2016.
- [7] A. Souahli, G. Clerc and H. Razik, "Detection and Diagnosis of faults in Induction Motor Using an Improved Artificial Ant Clustering Technique", *IEEE Transactions on Industrial Electronics* 60(9): 4053-4062, 2013.
- [8] L. Saidi, Jaouher Ben Ali, E. Bechhoefer, and M. E. H. Benbouzid, "Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR," *Applied Acoustics*, vol. 120, pp. 1-8, May 2017.
- [9] A. Rai, S.H. Upadhyay, "Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering," *Mechanical Systems and Signal Processing*, Volume 93, Pages 16-29, 1 September 2017.
- [10] C.T. Yiakopoulos, K.C. Gryllias, I.A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a K-means clustering approach," *Expert Systems with Applications*, vol. 38, pp. 2888-2911, 2011.
- [11] M. M. Etefagh, M. Ghaemi and M. Yazdaniyan Asr, "Bearing Fault Diagnosis using Hybrid Genetic Algorithm K-means Clustering," *Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on*, At Alberobello, June 2014.

I. Thématique initial de recherche :

Selon les dernière études, la sortie de la cinquième génération des réseaux (5G), va connaître la proposition de nouveaux concepts qui vont bouleverser les systèmes actuellement utilisés par les fournisseurs d'accès et les opérateurs mobiles. Ces nouveaux concepts proposés visent à améliorer les performances et la qualité de service fournie, par ces derniers, aux profit des clients, toute en diminuant les coûts. Ce bouleversement est due aux plusieurs facteurs: i) la grande évolution ainsi que la grande utilisation des équipements de télécommunications (smarts phone , tablets,..etc) ; ii) la propagation rapide de l'utilisation des applications intégrées par ces équipements, ce qui induits à une grande demande des ressources; iii) l'orientation vers une nouvelle tendance d'application dites intelligentes (Smart cities, Grid cities,) qui se base sur des nouvelles technologies comme IoT (Internet of Things) et les UAVs (Unmanned Aerial Vehicles); iv) l'apparition de nouveaux concepts pour architecturer les réseaux et optimiser leurs utilisations comme SDN(Software-Defined Networking), VNF(Virtualised Network Function), MEC(Mobile Edge Computing).

En outre, selon les dernières publications sur la standardisation, nous avons marqué l'apparition d'une nouvelle tendance dans la conception des réseaux. Cette dernière vise à aller au delà de l'architecture terrestre fixe du réseau vers la conception d'une plateforme mobile via l'intégration des objets mobile tel que les UAVs. En effet , notre contribution va se pencher précisément à proposer de nouveaux protocoles pour améliorer les performances du réseau 5G mobile en terme de:

- ✓ Minimisation de délai de transfert.
- ✓ Gestion efficace de QoS (Quality of Services) en minimisant la congestion dans le réseau.
- ✓ Assurer une meilleur QoE (Quality of Experience) tout en minimisant les couts, faciliter l'immigration et assurer la portabilité.

D'autre part, la contribution se focalise aussi dans l'intégration des approches SDN et VNF dans la conception de protocoles destinées à la gestion des réseau 5G .

II. Thématique cernée de la thématique initiale:

Les UAVs ou les drones ont connus , ces dernières années, un très grand succès. les UAVs ont été intégrés dans plusieurs types d'applications grâce à leurs, mobilités, connectivités aux réseaux. , la possibilité d'être télé-contrôlés ainsi être un support pour tout type d' équipements dotés d'une adresse IP. Parmi ces types d'applications nous avons, la livraison des courriers, la détection des catastrophes, la gestion routières, les villes intelligentes (smart cities) et pleines d'autres.

L'intégration des UAVs dans les différents types d'applications est face à plusieurs défi imposés d'une part par la nature constructive et fonctionnelle des UAVs (capacité de ressource limité, inter-connectivité non stable due au mobilité...etc). En effet , notre contribution va se pencher précisément à proposer de nouveaux protocoles pour améliorer les performances de fonctionnement des UAVs afin de faciliter son intégration dans les différents types d'applications.

III. Travaux réalisés

Mécanisme efficace assurant le calcul offline pour les UAVs:

Les UAVs sont envisagées d'être utilisés partout dans notre quotidien, intégrés dand les différents services et applications. Généralement les UAVs sont dotés des équipements IoT comme les capteurs sans fil, caméra,...etc. Pendant une mission spécifique les UAVs peuvent offrir plusieurs IoT services à valeur ajouté à condition qu'il soient équipés par les bons équipements. Cependant, ce genre de services de valeur ajouté consomme beaucoup de ressources en terme d'énergie ainsi en terme d'équipements, chose qui ne pourra pas être accomplis via un seul UAV. Comme solution a ce problème nous avons proposé un nouveau mécanisme qui assure la computation offline a travers un cluster d'UAVs. Ce mécanisme contient deux sous-mécanismes pour donner la possibilité aux UAVs d'accomplir les services de

valeur ajoutés d'une manière optimale. Le premier sous-mécanisme vise à faire une optimisation d'énergie des UAVs. Le second consiste à minimiser le temps de réponse des UAVs. Les deux sous-mécanismes sont modélisés via la programmation linéaire.

Cette solution est publiée dans l'édition 2017 de la conférence IEEE ICC sous le titre "**Efficient offloading mechanism for UAVs-based Value Added Services**".

IV. Travaux en cours

Actuellement nous travaillons sur :

1- L'extension journal de la solution précédente (**offloading mechanism for UAVs-based Value Added Services**). Dans cet axe nous sommes en phase de préparation d'un nouveau modèle plus générale pour garantir la computation offline dans un cluster d'UAVs offrant des services IoT de valeur ajoutée.

2- La préparation d'une nouvelle solution qui traite le problème de collisions dans un groupe d'UAVs. Cette solution propose deux algorithmes qui font respectivement la détection ainsi l'évitement des collisions dans un groupe d'UAVs. Cette solution est en phase de simulation.

Modèles pour la spécification des besoins des décideurs pour la conception des entrepôts de données

Waffa Setra

Directeur de thèse : Pr. Rachid Chalal

Résumé— Notre préoccupation de recherche est l'évolution des entrepôts de données. Sur la base de la synthèse de l'état de l'art du domaine, nous proposons de solutionner cette problématique dans un environnement de « Entrepôt de données NoSQL ». Un résumé succinct de nos travaux de recherche au cours de la 4^{ème} année doctorale est fourni.

I. INTRODUCTION

L'entrepôt de données (ED) avec ses principales caractéristiques : maintenir les données historisées, et fournir l'information utile pour la prise de décision, doit intégrer tous les changements influençant son efficacité. En effet, les sources de données qui alimentent l'ED sont sujettes à de multiples modifications, ainsi que les besoins des décideurs qui évoluent avec le temps.

De nos jours, on parle de plus en plus d'ED NoSQL. En effet, plusieurs contributions [1],[2], [3] se sont intéressées à cette problématique, et ont proposé d'adopter le NoSQL pour l'implémentation des ED, afin de bénéficier de ses avantages et performances.

II. PROBLEMATIQUE

Le cycle de vie de l'ED est composé de trois phases : la planification, la conception et implémentation et enfin la maintenance et évolution. C'est l'évolution de l'ED qui a comme objectif de garder ce dernier à jour et par conséquent éviter son obsolescence.

Les travaux de recherche [1],[2], [3] se sont concentrés principalement sur la deuxième phase du cycle de vie de l'ED (conception). Ces travaux n'ont pas fourni une solution pour garder l'ED à jour suivant les changements qui surviennent au cours du temps. Par conséquent la troisième phase du cycle de vie de l'ED n'a pas été considérée, et une fois l'ED se retrouve dans une situation d'évolution, cette tâche consommera énormément de temps et d'effort, du moment qu'elle n'a pas été planifiée.

Notre principale préoccupation de recherche est de prendre en charge l'aspect évolutif de l'ED dans un environnement NoSQL, indépendamment de son approche de conception. En effet, un ensemble d'hypothèses à respecter est formulé.

Les principales questions posées est qui sont à résoudre sont:

- 1- Sachant qu'il existe plusieurs implémentation NoSQL, quelle est la plus adéquate pour prendre en charge l'évolution.
- 2- Les approches d'évolution des ED existantes sont-elles applicables à un ED NoSQL.

- 3- Si non, comment solutionner ce problème.

III. TRAVAUX REALISES

Au cours de notre 4^{ème} année doctorale, nous avons défini les grandes lignes de notre proposition en réponse aux questions précédentes :

- Choisir l'implémentation NoSQL à Adopter qui soit la plus adéquate pour prendre en charge l'évolution de l'ED,
- Choisir la catégorie d'évolution à emprunter dans le cas de l'implémentation NoSQL choisie,
- Adaptation, application et test de certains travaux d'évolution d'ED existants sur l'environnement défini.

En résumé, nous sommes dans la phase implémentation de notre approche, cependant le manque de performance des machines nous a été perturbant.

IV. PERSPECTIVES

Ce qui reste à faire pour le futur proche :

- Finaliser la phase des tests et implémentation,
- Publier les résultats,
- Rédiger la thèse.

V. CONCLUSION

En conclusion, nous jugeons qu'il serait très bénéfique si nous bénéficierons d'une assistance ou aide pour la rédaction d'articles en anglais, des chapitres de la thèse.

Références

- [1] M. Chevalier, M. El Malki, A. Kopliku, O. Teste, and R. Tournier, "Entrepôts de données multidimensionnelles NoSQL," in *EDA*, 2015, pp. 161-176.
- [2] M. Chevalier, M. El Malki, A. Kopliku, O. Teste, and R. Tournier, "Document-Oriented Data Warehouses: Complex Hierarchies and Summarizability," vol. 397, pp. 671-683, 2017.
- [3] K. Dehdouh, F. Bentayeb, O. Boussaid, and N. Kabachi, "Using the column oriented NoSQL model for implementing big data warehouses," in *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)*, 2015, p. 469.

Une approche d'analyse de sentiment de texte en dialecte arabe basée sur la traduction automatique. Application à l'Algérien.

Imane GUELLIL

Directeurs de thèse : Faïçal AZOUAOU. MCA ESI.

Équipe : (LMCS)

Résumé— Les pages Facebook Algériennes contiennent des commentaires de la forme : 1) « ما نميناجيش لقرابة هلكتي و أنا ما قديتش » voulant dire « je ne vais pas ménager. Les études m'ont détruit et je n'ai pas pu ». 2) « rani dayardjezy 3jabtni mlih » voulant dire « j'ai mis djezzy elle me plaît bien ». 3) « oreedoo نستعرف بيكم يا oreedoo » voulant dire « on vous tire chapeau oreedoo ». Les trois commentaires expriment le sentiment de l'utilisateur vis-à-vis d'un sujet particulier. Néanmoins au sein du premier commentaire l'utilisateur fait appel aux lettres arabes pour s'exprimer. Au sein du deuxième, il utilise l'arabizi (qui représente un mélange entre lettres arabes et latines. Au sein du Troisième, l'utilisateur alterne entre lettres arabes et arabizi. Nous constatons que le premier message est négatif alors que les deux autres sont positifs. Le but de cette thèse est donc de proposer et de combiner plusieurs approches permettant de faire cette analyse (classifier les messages Algérien en positifs et négatifs) de manière automatique.

I. INTRODUCTION

L'arabe est une langue riche et complexe parlée par plus de 400 millions de locuteurs. Cette langue se trouve cependant dans un état de diglossie dans les pays où elle est utilisée vu qu'elle coexiste avec 22 dialectes regroupé par la plupart des chercheurs en six catégories : 1) l'égyptien, 2) Le magrébin, 3) l'irakien, 4) le levantin, 5) Le saoudien et 6) Les autres pour le soudan et certain pays du sud de l'Afrique [1, 2]. L'émergence des médias sociaux a suscité un intérêt particulier du traitement de l'arabe et ces dialectes. Parmi les pistes de recherches les plus étudiées; l'analyse de sentiment, qui permet de déterminer si un message donné est plutôt positif ou négatif. Plusieurs travaux ont été menés sur l'analyse de sentiment de l'arabe et de ses dialectes. Les approches utilisées au sein de ces travaux peuvent être supervisée, non-supervisée ou hybride [3, 4]. Par exemple, au sein du travail [5], les auteurs étudient trois techniques de construction de lexique dont une technique manuelle et deux autres automatiques. En plus de cela, un outil d'analyse de sentiment a été développé. Ce travail a été mené sur l'arabe standard et ses dialectes. Nous citons également le travail dans [6], qui propose une méthode hybride basée sur le lexique et le modèle bayésien naïf en même temps. La méthode proposée contient une phase de prétraitement (transformation, normalisation, segmentation,...).

L'approche à base de lexique est exécutée pour remplacer quelques mots avec leurs synonymes. La tâche de classification est effectuée en faisant appel au modèle bayésien naïf pour classer le sentiment en positif et négatif.

Nous n'avons cependant pu identifier aucun travail spécifique à l'analyse de sentiment du dialecte algérien. Le dialecte Algérien est un dialecte magrébin parlé par plus de 40 million de locuteurs (ce qui représente plus de 10% de la population s'exprimant en arabe). Le but de cette thèse est donc de proposer une approche globale permettant l'analyse d'un ensemble de messages donné en Dialecte Algérien

II. Problématique

Sur le média social Facebook, en particulier sur les pages dédiées aux Algériens, nous trouvons des commentaires tels que : 1) « ما نميناجيش لقرابة هلكتي و أنا ما قديتش » voulant dire « je ne vais pas ménager. Les études m'ont causé détruit et je n'ai pas pu ». 2) « rani dayardjezy 3jabtni mlih » voulant dire « j'ai mis djezzy elle me plaît bien ». 3) « oreedoo نستعرف بيكم يا oreedoo » voulant dire « on vous tire chapeau oreedoo ». Nous pouvons constater que les utilisateurs Algériens des médias sociaux font appel à plusieurs scripts pour l'écriture : L'Arabe, l'Arabizi ou encore un mélange des deux. Donc la première problématique à laquelle est confrontée l'analyse de sentiment du dialecte Algérien est la normalisation entre ces deux scripts. Sans oublier les problématiques partagées avec la langue arabe telles que : 1) La dérivation (plusieurs schémas pour le même radical). 2) L'inflexion (variation d'un mot selon son temps ; son genre, son nombre,) et surtout 3) L'agglutination (plusieurs mots attachés les uns aux autres. Il ne faut pas oublier, qu'en dialecte Algérien, même la négation peut être agglutinée. Quel est donc le modèle, l'approche, la méthode et la technique la plus appropriée à utiliser pour analyser le sentiment de message ayant de telles caractéristiques ?

III. TRAVAUX REALISES

Avant de commencer ce doctorat, un état de l'art sur la fouille des médias sociaux et plus particulièrement l'analyse de sentiment a été dressé. Ce travail a fait l'objet d'une publication dans une conférence internationale ISPS 2015 [7].

Au cours de notre première année, nous nous sommes d'abord concentrées sur l'aspect morphologique du dialecte Algérien où nous avons proposé un analyseur syntaxique de ce dialecte. Ce dernier peut donc séparer le radical de ces préfixes et suffixes. Ce travail a fait l'objet d'une publication à la rencontre internationale des jeunes chercheurs RJCIA 2016 [8]. Par la suite, nous avons exploité le travail précédemment établi sur l'analyse morphologique pour pouvoir distinguer entre le Dialecte

Algérien, l'arabe standard et le Français au sein des médias sociaux. Ce travail a également fait l'objet d'une publication dans un workshop se déroulant au sein la conférence internationale CSE2016 [9].

Au cours de notre deuxième année, nous nous sommes concentrés sur l'aspect de l'Arabizi. Nous nous sommes donc penchés sur la translittération (qui consiste à standardiser un message écrit en Arabizi en arabe. Nous aurions ainsi tous nos messages écrits au sein du même script. Ce travail a fait l'objet d'une publication au sein du workshop SocialMT2017 hébergé au sein de la conférence EAMT2017 [10]. Nous avons ensuite exploité ce travail de translittération pour proposer une approche de traduction des messages après leur translittération, pour démontrer la hausse des résultats grâce à la translittération. Ce travail a fait l'objet d'une publication au sein du workshop WiNLP2017 (workshop de la conférence ACL2017) [11]. Nous avons également proposé au cours de cette année une méthode de construction et d'enrichissement d'un dictionnaire en Arabizi. Ce travail a également fait l'objet d'une publication au workshop WiNLP2017 [12].

IV. PERSPECTIVES

Nous travaillons actuellement sur la construction d'un lexique de sentiment du dialecte Algérien (en arabe et en arabizi). Nous travaillons également sur l'exploitation de cette ressource au sein d'un Algorithme (en cours de développement) basé sur les lexiques construits et permettant l'analyse du message en dialecte Algérien. Dans nos prochains travaux nous allons nous intéresser à l'exploitation de la translittération et de la traduction dans un but d'analyse de sentiment.

V. CONCLUSION

L'analyse de sentiment est une thématique de recherche très étudiée ces dernières années. Elle permet de déterminer l'orientation sémantique (positive, négative ou neutre) d'un texte donné. Vu la richesse morphologique de l'arabe et de ses dialectes, pouvoir déterminer cette orientation est très compliqué. Nous nous concentrons donc dans ce doctorat sur la proposition d'une approche d'Analyse de sentiment du dialecte Algérien. Pour ce faire, nous avons commencé par l'analyse morphologique de ce dialecte, translittération de l'arabizi et sur l'identification du dialecte Algérien. Nous visons à court terme la proposition d'un lexique de sentiment du dialecte Algérien ainsi que la proposition d'un Algorithme d'Analyse de sentiment basé sur ce lexique.

Nous envisageons cependant à moyen terme d'analyser les sentiments des messages écrits en dialecte Algérien avant et après la translittération, avant et après la traduction et avant et après l'enrichissement de notre lexique. Notons que nous travaillons également sur un Algorithme d'enrichissement d'un lexique de sentiment à l'aide du deep learning.

References

- [1] O. F. Zaidan and C. Callison-Burch, "Arabic dialect identification," *Computational Linguistics*, vol. 40, pp. 171-202, 2014.
- [2] F. Sadat, F. Kazemi, and A. Farzindar, "Automatic identification of arabic language varieties and dialects in social media," *Proceedings of SocialNLP*, p. 22, 2014.
- [3] M. Biltawi, W. Etaiwi, S. Tedmori, A. Hudaib, and A. Awajan, "Sentiment classification techniques for Arabic language: A survey," in *Information and Communication Systems (ICICS), 2016 7th International Conference on*, 2016, pp. 339-346.
- [4] A. Assiri, A. Emam, and H. Aldossari, "Arabic sentiment analysis: A survey," *International Journal of Advanced Computer Science and Applications*, vol. 6, pp. 75-85, 2015.
- [5] N. Abdulla, S. Mohammed, M. Al-Ayyoub, and M. Al-Kabi, "Automatic lexicon construction for arabic sentiment analysis," in *Future Internet of Things and Cloud (FiCloud), 2014 International Conference on*, 2014, pp. 547-552.
- [6] K. Khalifa and N. Omar, "A hybrid method using lexicon-based approach and naive Bayes classifier for Arabic opinion question answering," *Journal of Computer Science*, vol. 10, p. 1961, 2014.
- [7] I. Guellil and K. Boukhalfa, "Social big data mining: A survey focused on opinion mining and sentiments analysis," in *Programming and Systems (ISPS), 2015 12th International Symposium on*, 2015, pp. 1-10.
- [8] I. GUELLIL and F. Azouaou, "ASDA : Analyseur Syntaxique du Dialecte Algérien dans un but d'analyse sémantique," in *Conférence Nationale d'Intelligence Artificielle Année 2016*, 2016, pp. 87-94.
- [9] I. Guellil and F. Azouaou, "Arabic Dialect Identification with an Unsupervised Learning (Based on a Lexicon). Application Case: ALGERIAN Dialect," in *Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), 2016 IEEE Intl Conference on*, 2016, pp. 724-731.
- [10] I. Guellil, F. Azouaou, M. Abbas, and S. Fatiha, "Arabizi transliteration of Algerian Arabic dialect into Modern Standard Arabic," in *Social MT 2017/First workshop on Social Media and User Generated Content Machine Translation*, 2017.
- [11] I. Guellil, F. Azouaou, and M. Abbas, "Comparison between Neural and Statistical translation after transliteration of Algerian Arabic Dialect," in *WiNLP: Women & Underrepresented Minorities in Natural Language Processing (co-located with ACL 2017)*, 2017.
- [12] I. GUELLIL and A. Faical, "Bilingual Lexicon for Algerian Arabic Dialect Treatment in Social Media," in *WiNLP: Women & Underrepresent*

Towards New Approach for the Enhancement of Adaptive Video Streaming

Oussama El Marai

Pr. Mouloud Koudil; Dr, Mohamed Menacer:

Équipe : (CoDesign, Optimisation, MSI, EIAH, Image, SURES)

Abstract—Many studies have shown that the Dynamic Adaptive Streaming over HTTP scheme is limited in achieving efficiency, fairness and stability. To this aim, we proposed a client-server cooperation-based approach to achieve the three objectives. We also proposed a mathematical formulation of the problem. The proposed approach is evaluated through extensive simulations and compared against that of notable algorithms, such as the FESTIVE scheme. The results show that the proposed approach defines a promising way in enhancing the aforementioned objectives.

I. INTRODUCTION

Video applications usually manipulate huge content that require large bandwidths and dedicated protocols and methods to efficiently deliver the video content. A highly important video streaming approach that has attracted many researchers and practitioners in the last years is the Dynamic Adaptive Streaming over HTTP (DASH) scheme. Despite its wide popularity, previous studies showed that DASH faces several challenges relevant to stability, fairness and efficiency, particularly when multiple clients compete for the same link, creating a bottleneck link scenario.

To achieve a stable system, the client should detect the short term bandwidth variations and accordingly avoid frequent short-term bit-rate switches. The efficiency means that the client should watch the highest possible quality with respect to its available throughput. Also, the system must ensure fairness among the connected clients. Thus, the quality of experience (QoE) perceived by the clients having equivalent bandwidth capacities should be close to each other. All these challenges make the design and development of a robust DASH system highly difficult since they are challenging and conflicting each other. Moreover, an ideal DASH system should allow quick convergence of the clients to the fair share and ensure smoothness when increasing or decreasing the quality. Effectively, frame freezes shall be avoided when switching to the next higher/lower level and only the right number of chunks shall be requested from each level.

In this report, we will briefly summarize our works related to the enhancements of adaptive video streaming during the previous years.

II. PROBLEM DEFINITION

The DASH technology aims at provisioning a high QoE to the users by dynamically adapting the bit-rate of the played video to the clients' conditions (e.g., bandwidth fluctuations and device capabilities). This can be achieved by storing multiple versions of the same video at the server side where each video is encoded into a set of m discrete representation levels with different encoding rates $L = \{l_0, l_1, \dots, l_{m-1}\}$, with $l_0 < l_1 < \dots < l_{m-1}$. Each representation level l_j , $0 \leq j < m$, is chopped into n small segments of fixed seconds called chunks. We denote by c_{ij} the i -th chunk from the j -th representation

level. The server is connected to the internet with a W bandwidth link which gives an upper-bound value on how fast the server's connection can possibly transmit data. The server's link capacity, also called bottleneck link, is shared among N competing clients where each client k has its own end-to-end estimated link capacity T_i after downloading the i -th chunk. The downloaded chunks are buffered into a space memory of max seconds maximum. The rate of consumption from the buffer is constant with one second of the buffered video time each real-time second. Based on both client's estimated bandwidth and buffer state, the client selects the adequate representation level j to request. When the buffer goes empty, the end user will experience video playback freezes which is the main problem that should be avoided as much as possible even if the video bit-rate is decreased. For this aim, it might be relevant to define a minimum buffer threshold \min which serves as an alarm for the client to decrease the played representation level once attempted. On the other hand, if the buffer is full, the download process goes idle (OFF period) for a calculated period until enough space is available for at least one chunk. The OFF periods cause alternations in the observed throughput which ultimately impact the perceived QoE.

III. ACHIEVED WORKS

A. Problem Formulation

We formulated the defined problem as follows:

$$\text{Min } Z = \sum_{i \in C} \sum_{j \in L} p_{ij} x_{ij}$$

Subject to:

$$\sum_{i \in C} \sum_{j \in L} l_j x_{ij} \leq B \dots (1)$$

$$\forall i \in C, \sum_{j \in L} x_{ij} = 1 \dots (2)$$

$$\forall i \in C, \sum_j l_j x_{ij} \leq r_i \dots (3)$$

$$\forall i \in C \setminus n - 1, \sum_{j \in L} x_{ij} - \sum_{j \in L} x_{(i+1)j} \geq 0 \dots (4)$$

Where Z_1 is the objective function that maximizes the clients' efficiency and Z_2 is the objective function that ensures fairness between the different clients.

x_{ij} : Binary variables. $x_{ij} \in \{0,1\}$.

B : Shared Bandwidth.

C : Set of clients. $i \in \{0,1,2, \dots, n - 1\}$.

L : Set of different levels. $j \in \{0,1,2, \dots, m - 1\}$.

r_i : Requested level of client i .

p_{ij} : Represents the score of Client i when it is moved from family f_j to f_{j-1} .

The matrix p defines a priority, by attributing scores at each representation level, among the different clients when a level decreasing becomes necessary. As our objective function aims at minimizing the total sum of p values, the scores are attributed in a way that the clients having requested high representation levels get lower values.

To evaluate the performance of the proposed model, we conducted many simulations using the NS3 simulator, and compared the results to FESTIVE and ESTC. The proposed model is implemented using the Optimization Programming Language (OPL) and resolved with the CPLEX optimizer from IBM.

B. Proposed ESTC solution

Our proposed approach [1] to achieve a fair system that delivers the highest possible video quality while remaining stable is to delegate to each side the appropriate objective based on the available information. For the client-side, the previously played chunks, the buffer state and the estimated bandwidth are available. So, both stability and efficiency are tuned there. Information regarding the number of connected clients, the amount of shared bandwidth are however known by the server which make it the best to ensure fairness between clients. The overall architecture of ESTC is depicted in Fig. 1.

1) Client-side Algorithm

The roles of the client’s controllers consists of maximizing the perceived video quality, minimizing the oscillations and ensuring smoothness in the video playback upon changes in the streaming bit-rate, while maintaining the buffer above a predefined threshold. The output of the client controllers is the final decision on the next chunk’s bit-rate. This decision, along with the client’s estimated bandwidth capacity, is communicated to the server.

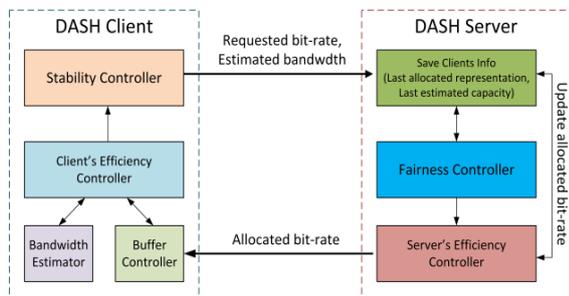


Fig. 1. Global architecture of the proposed ESTC scheme.

2) Server-side Algorithm

At the server side, the shared bandwidth W as well as the number of connected clients N , their current downloading bit-rate and last smoothed throughput are known (i.e., since each client sends these information to the server in the HTTP GET requests). This places the server in a better position to efficiently control and fairly distribute the shared bandwidth among the connecting clients. To this end, the server uses the information received from the different clients (e.g., selected bit-rate, smoothed throughput) to make a fair decision on which bit-rates should be allocated to the clients. Furthermore, it exploits the remaining throughput of the clients who do not fully use their allocated bandwidth to enhance the overall system efficiency. The benefit of this strategy is to maximize the bandwidth utilization and, at the same time, it allows a quick and smooth convergence of the different clients, depending on

their bandwidth capacity, to the fair share. It also ensures stable video streaming to the different clients by prohibiting others from exploiting the remaining bandwidth randomly due to OFF periods. Note that the server can supply a client with either the requested or lower bit-rate, and does not allocate higher than the requested bit-rate.

C. Development Concepts of a Media Management System Dedicated to Quran and Islamic Education

In this work [2][3], we proposed a media streaming web-based system dedicated for Holy Quran and Islamic education. It is composed of 3 main services: Video-on-Demand, Scheduled Program and Live Events. It combines both traditional streaming and DASH technology for delivering media to PC-based and Handhelds clients respectively. For PCs where the bandwidth is generally quite constant and the performance of the device like CPU and Memory are sufficient, in this case WMS (Windows Media Services) is used that implements traditional streaming. However, handheld devices like smartphones and tablets are characterized by their frequent mobility that causes bandwidth fluctuations. In addition, the performance of these devices is lower compared to PCs; in this case DASH technology is used.

IV. PROSPECTS

Currently, we are working on a new DASH-based solution to improve the video streaming QoE by exploiting the powerful features of SDN networks using ONOS platform. Depending on the clients’ capacity, the solution consists of ensuring a per-client minimum QoE.

V. CONCLUSION

During the previous years, we developed a media streaming system QMS dedicated to learning Quran and its Sciences. It consists of three main services: Video-on-Demand (VoD), WebTV with scheduler, and Live streaming. From this work, we published two conference papers [3], [4]. After that, we proposed ESTC solution [1] to build a robust DASH-based system aiming to achieve the following objectives: high utilization of network resources, stability, fairness, and a short time of convergence to the fair share while avoiding buffering. The paper is currently in press in IEEE Transactions on Broadcasting. A mathematical formulation of the previous work was submitted to Globecom’17 conference and it was recently accepted. Recently, we started working on a SDN-based architecture with Prof. Taleb Tarik (Aalto University - Finland) to improve the per-client QoE in DASH-based system. In parallel, we are also writing the thesis which is approximately 35% done.

Publications

- [1] Elmarai. O, Taleb. T, Menacer. M, Koudil. M. “On Improving Video Streaming Efficiency, Fairness, Stability & Convergence Time through Client-Server Cooperation”. In Press in IEEE Trans. on Broadcasting, 2017.
- [2] Elmarai. O, Taleb. T. “Online Server-side Optimization Approach for Improving QoE of DASH Clients”. Accepted in Globecom’17 conference, 2017.
- [3] Elmarai. O, Menacer. M, and Arbaoui. A. "Development concepts of a media management system dedicated to Quran and Islamic education." 2013 International Conference on Current Trends in Information Technology (CTIT). IEEE, 2013.
- [4] Menacer. M, Elmarai. O, and Arbaoui. A. "Design and architecture of a video streaming system dedicated to Quran related multimedia resources." Computer Applications & Research (WSCAR), 2014 World Symposium on. IEEE, 2014.

Un framework pour la fouille visuelle des données basé sur l'architecture Peer-to-Peer & la technologie des jeux vidéo online

Abdelkrim Hamadi

Directeurs de thèse : Fahima Nader

Équipe : Optimisation

Résumé— Le développement phénoménal du matériel informatique permet aux systèmes informatiques de stocker des masses de données gigantesques, cependant l'exploitation et l'analyse de ces volumes de données devient un travail fastidieux.

La fouille visuelle de données (VDM : Visual Data Mining) peut bénéficier de la puissance et des performances de l'architecture Peer-to-Peer (P2P) qui offre des ressources de calcul et de stockages impressionnantes et aussi de la technologie des jeux vidéo online. C'est dans ce contexte que se situe notre travail, on va décrire puis implémenter un framework appelé Visual Data Mining Distributed using Game (VDMDG) [1] qui permettra d'exploiter des algorithmes VDM dans un environnement P2P en utilisant les concepts des jeux vidéo online.

I. INTRODUCTION

La fouille des données (DM : Data Mining) a connu un essor important depuis les années 90 et cela depuis la mise au point de nouveaux algorithmes performants permettant de traiter un grand volume de données. Le DM a réussi à prouver son utilité dans un certain nombre d'applications, et cette expérience a conduit à concevoir des normes appelées méthodologie CRISP-DM [2][3]. Cette méthodologie identifie cinq phases pour chaque tâche de fouille de données complexes : la spécification des tâches, la compréhension des données, la préparation des données, la modélisation et le déploiement. L'expert de la fouille de données a besoin d'outils permettant d'évaluer rapidement le type des relations cachées dans les données fournies. Une telle information peut l'aider à identifier les sous-ensembles intéressants dans les données, de fournir rapidement des données claires et pertinentes au client.

Les méthodes de fouille visuelle de données (VDM : Visual Data Mining) tentent de résoudre les problèmes d'interprétation et d'interaction dans les processus de découverte de connaissances en faisant appel à des visualisations dynamiques et à des requêtes graphiques sur les données et connaissances représentées [4][5][6]. Face à la diversité des besoins en informations, de nombreuses techniques de visualisation et d'interaction ont émergé durant ces dernières années. A titre d'exemples classiques, nous pouvons citer les visages de

Chernoff [7] qui représentent des données sous la forme d'icônes en s'appuyant sur le fait que l'esprit humain analyse facilement les ressemblances et différences entre visages. Nous pouvons citer également les "scatter plots" [8] qui permettent d'obtenir des vues multiples sur les données et d'observer les données à l'aide de techniques graphiques comme le "brushing" qui donne la possibilité de sélectionner des données dans une vue tout en soulignant ces mêmes données dans les autres vues.

II. PROBLEMATIQUE

De nombreux travaux ont été menés sur la problématique de la visualisation d'information et de la fouille visuelle de données. Il s'agit en effet d'un problème de plus en plus important. La quantité d'information ne cesse de croître de jours en jours. Par conséquent, il est impossible pour les experts de se passer d'outils de recherche automatiques. Ces données constituent une source d'information importante, il est donc nécessaire de penser à concevoir des méthodes efficaces permettant d'extraire, de manipuler et de mettre en forme les informations qu'elles peuvent contenir.

Quatre problèmes majeurs se présentent :

- Traiter de grands volumes de données.
- Traiter tous types de données.
- Permettre l'utilisation de différentes techniques de fouille de données via ce Framework.
- Optimisation du temps de réponse.

III. TRAVAUX REALISES

Le nouveau framework proposé VDMDG se base sur une topologie réseau hybride P2P ou les ordinateurs peuvent être des clients ou bien des serveurs. Le framework VDMDG offre les caractéristiques suivantes :

- Une très grande disponibilité de service.
- Une transparence dans la répartition de la charge de calcul et des données.
- Une redondance des données.

L'architecture du framework VDMDG repose sur les nœuds suivants :

- « DatabasePeer » : Ce nœud est responsable sur l'ensemble des accès aux bases de données. Le nœud « DatabasePeer » est le seul interlocuteur entre le groupe des nœuds « VDM Peer group » et les serveurs de base de données, ces

- nœuds reçoivent des requêtes du groupe des nœuds «VDM Peer group » et communiquent ces demandes aux serveurs des bases des données en utilisant le connecteur « DB Connector » ainsi que les résultats des requêtes. Les bases de données (tables et indexes) peuvent être fragmenté sur plusieurs serveurs. Les nœuds « DatabasePeer » sont regroupés dans le groupe « DB Peer Group ».
 - « OverviewPeer » : Ce type de nœud est responsable de communiquer à l'utilisateur une vue globale sur les données.
 - « FilterPeer » : Ce type de nœud est chargé d'appliquer des filtres sur les données.
 - « DetailPeer » : Ce nœud se consacre à l'exécution des requêtes à la demande de détails sur les données.
 - « ZoomPeer » : Ce nœud donne plus de détails sur une sélection de données.
 - « BusinessPeer » : Ce nœud joue le rôle d'un tampon entre les noeuds « GUIClientPeer » et le groupe des nœuds « VDM Peer Group » (OverviewPeer, FilterPeer, DetailPeer and ZoomPeer). Ce nœud reçoit simplement des paquets puis les transfère vers le groupe de noeuds « VDM Peer Group ».
 - « GatheringPeer » : Ce nœud collecte les données résultantes du groupe de nœuds exécutant les algorithmes de la fouille de données « DM Peer Group » et sauvegarde les résultats de ces algorithmes dans un deuxième groupe de base de données « DB Peer Group », ces données seront par la suite exploiter par les nœuds « Business Peer ».
 - « Datamining Peer » : Ces nœuds sont chargés d'exécuter les programmes de la fouille de données.
 - « GUI ClientPeer » : Ces nœuds sont responsables de l'affichage des résultats à l'écran de l'utilisateur. Ces nœuds sont des applications clientes qui interagissent avec les nœuds « BusinessPeer ».
- [1] A. Hamadi, F. Nader , "Visual Data Mining Framwork based on Peer-toPeer Architecture," Proceedings of the Ninth International Conference on Digital Society ICDS 2015, Lisbon, Portugal, vol. 1, pp. 14,February 2015. Lasse Berntzen, HBV, Norway publisher, ISBN: 978-1-61208-381-0
 - [2] W. S. Cleveland. Visualizing Data. Hobart Press, Summit, New Jersey, U.S.A., 1993.
 - [3] Ben Shneiderman. The eyes have it : A task by data type taxonomy for information visualizations. In IEEE Visual Languages, number UMCPCSD CS-TR-3665, pages 336–343, College Park, Maryland 20742, U.S.A., 1996.
 - [4] Pak Chung Wong et R. Daniel Bergeron. 30 years of multidimensional multivariate visualization. In Scientific Visualization — Overviews Methodologies and Techniques, pages 3–33. IEEE Computer Society Press, Los Alamitos, CA, 1997.
 - [5] Yijun Chen, Adbolreza Abhari. "A Service-Oriented Distributed Data Mining Prototype Based on JDM », Proceedings of the 2008 Spring simulation multiconference.
 - [6] Java Data Mining API 2.0, JSR 247. <http://www.jcp.org/en/jsr/detail?id=247>
 - [7] Joseph D. Gradecki(2003). « Mastering JXTA : Building Java Peer-to- Peer Applications ». Wiley Publishing, Inc., ISBN : 0-471-250848.pp15-38.

IV. PRESPECTIVES

Nous planifions pour l'année prochaine de faire les actions suivantes :

- Faire des tests avec le simulateur qui est en cours de finalisation.
- Publier les résultats de nos tests dans un article.
- Terminer la rédaction de la thèse.

V. CONCLUSION

Le prototype du framework VDMDG développé autour de l'architecture P2P et de la technologie des jeux vidéo online offre un très bon temps de réponse et une très grande disponibilité de service.

Sécurisation des dispositifs biométriques mobiles

Hachemi Nabil Dellys

Mouloud Koudil et Karima Benatchba

Équipe : SURES.

Résumé— Cette thèse a pour objectif initial l'étude de la sécurisation des dispositifs biométriques mobiles. Plusieurs méthodes de sécurisation des systèmes biométriques ont été étudiées parmi les crypto-systèmes biométriques et la biométrie résiliable. Nous nous sommes concentrés sur la méthode du Fuzzy Vault qui est une technique de cryptage des templates biométriques tolérante à un certain seuil d'erreur. Cette méthode utilise une clé secrète en plus des données biométriques pour crypter le template.

Mot Clés : Biométrie, Sécurisation biométrique, Crypto-Système biométrique, Fuzzy Vault, Chaff-point generation.

I. INTRODUCTION

La biométrie est l'une des trois approches de contrôle d'accès, avec la connaissance secrète et la possession de carte d'accès. Plusieurs systèmes de contrôle d'accès ont vu le jour, utilisant une panoplie de modalités biométriques (Empreinte digitale, visage, iris, voix etc.). Les recherches ne cessent d'évoluer pour optimiser les performances de traitement et de réduire les taux d'erreurs.

Cette approche présente, cependant, un grand inconvénient qui est la possibilité d'atteinte à la vie privée. Le caractère quasi permanent des modalités biométriques offre la possibilité à un imposteur de compromettre tout un système s'il a accès à ses templates biométriques, dans le cas où ces dernières ne sont pas sécurisées. De plus, la biométrie permet de connaître les déplacements et les agissements de chaque personne. Le visage, l'iris ou à un degré moindre la technique de marche peuvent être capturés à distance et sans la coopération de l'individu, ce qui rend le suivi de personnes une opération assez simple.

La sécurisation des systèmes biométriques contre les attaques est devenue une nécessité pour la pérennité de cette approche. L'utilisation des techniques de cryptographie classiques s'avère impossible, car elles nécessitent des données quasi-exactes pour reproduire les mêmes informations cryptées. Les captures d'une même modalité biométrique sont différentes d'une capture à l'autre, elles ne permettent pas d'effectuer ce type de cryptage. De ce fait, un nouveau domaine de recherche a vu le jour : la sécurisation des systèmes biométriques.

II. PROBLEMATIQUE

Pour pouvoir concevoir des méthodes de sécurisation, il faut d'abord cerner les menaces. Identifier chaque menace à part n'offre pas une vue globale pour une solution

satisfaisante, car les attaquants trouvent toujours de nouvelles stratégies pour déjouer le système.

Des modèles de menaces sur les dispositifs biométriques ont été proposés pour donner une idée générale et/ou détaillée des menaces qu'encourent les systèmes biométriques du point de vue des vulnérabilités, du type d'attaques, des personnes qui menacent ces systèmes etc.

Plusieurs méthodes de sécurisation ont été proposées pour contrer ces menaces. Ces dernières ont été réparties en famille : Les crypto-systèmes biométriques, la biométrie résiliable et les combinaisons entre biométrie et cryptographie classique. A cela s'ajoutent des techniques de sécurisation des communications et vérification continue à l'aide de la biométrie.

III. TRAVAUX REALISÉS

Le travail effectué auparavant concerné les attaques et les modèles d'attaques sur les systèmes et dispositifs biométriques qui a abouti sur un article sur la comparaison entre les modèles d'attaque existant soumis en mois d'octobre 2012 à la conférence CaSon 2012 (Information Assurance and Security) a été sélectionné puis accepté pour une version longue dans la revue JIAS (Journal of Information Assurance and Security). Nous avons entamé dans le cadre de ce doctorat une étude sur les méthodes de sécurisation des systèmes biométriques en générale et des dispositifs biométriques mobiles en particulier.

En premier lieu nous avons fait une étude générale sur les techniques de sécurisation des systèmes biométriques. De cette étude nous avons ressorti trois grandes familles de méthodes: Les crypto-systèmes biométriques, la biométrie résiliable, et la sécurisation des communications utilisant les données biométriques. Chaque famille se décline vers plusieurs techniques. à titre d'exemple, les principales techniques appartenant à la famille des crypto-systèmes biométriques sont les techniques du Fuzzy vault, du Fuzzy commitment, du Fuzzy sketches et du Fuzzy extractor.

Parmi les méthodes de sécurisation basées sur les crypto-systèmes biométriques on trouve la méthode du Fuzzy Vault. Le Fuzzy Vault est une technique de sécurisation tolérante aux erreurs qui a été proposée pour la première fois en 2002 par Juels et Sudan, afin de pouvoir décrypter des informations même avec un certain seuil d'erreur entre les données de cryptage et de décryptage.

La technique du Fuzzy Vault a été adaptée à la biométrie. A l'aide d'une clé secrète, les caractéristiques biométriques sont réévaluées et mélangées à des points

trompeurs afin de rendre difficile la détection des points authentique.

D'après notre étude, nous avons conclu que le Fuzzy vault été plus intéressent pour les dispositifs biométriques mobiles, du fait qu'elle utilise une clé secrète en plus des templates biométriques. Etant donné que la sécurisation des communications et authentification biométrique étant ainsi que le stockage des template sont très important dans les dispositifs mobiles, l'utilisation de clés est très recommandée. L'acquisition illégale de la modalité ne compromettra pas le template biométrique.

Une étude détaillée des principaux travaux sur le Fuzzy vault a été effectuée. Une comparaison théorique des étapes de ce dernier nous a permis de rédiger un article [5] à la conférence IVIC'15 (International Visual Informatics Conference 2015).

Le PFE proposé en 2013 n'ayant pas aboutis a un une application exploitable, un autre PFE a été proposé pour une plateforme de tests modulable implémentant les étapes du Fuzzy Vault. Les étudiants ayant fait un très bon travail cette fois-ci, ce qui a aboutis sur une plateforme ou les principales propositions faite dans le Fuzzy Vault ont été implémentés.

Cette plateforme nous permet d'effectuer des tests et comparaison sur la génération des Chaff-points, ce qui a aboutis sur un article [3] à la conférence SoCPar'15 (Soft Computing and Pattern Recognition 2015). Cet article a été choisi parmi les meilleur et a fait l'objet d'un article de journal [2]. Une autre série de tests sur l'étape d'alignement a été effectuées et a fait l'objet d'un article en cours de soumission.

L'étude sur la génération des Chaff-points nous a permis de proposer une nouvelle méthode basée sur le problème du sac à dos et qui a fait l'objet d'une publication [1] à la conférence HIS'16 (Hybrid Intelligent System).

Enfin, Une étude complète sur les étapes les plus critiques du Fuzzy Vault à été effectuée. Cette étude c'est porté sur les étapes de génération des Chaff-points et l'alignement des points. Les résultats ont démontré que le meilleur compromis entre performance de reconnaissance, sécurité et temps d'exécution est l'utilisation d'une représentation composite avec la génération de Chaff-points à l'aide de la méthode des carrées. Cette étude à fait l'objet d'une publication au journal « Journal of Information and Communication Technology ».

IV. PERSPECTIVES

Comme cité précédemment, des tests comparatifs sont en cours de finalisation concernant les principales contributions dans chaque étape du Fuzzy Vault.

La thèse étant en cours de finalisation, la rédaction du mémoire devra commencer incessamment, un article de

revue est aussi en cours regroupant les résultats détaillés des tests effectués.

Des travaux sur l'utilisation des méthodes bio-inspiré sont intéressantes, des travaux connexes sont effectuée pour le moment en attendant de les testé dans le domaine de la sécurisation. La simulation de la cryptographie quantique pour la sécurisation des clés secrètes à l'aide de la biométrie sont des perspectives intéressantes.

V. CONCLUSION

L'étude bibliographique de la sécurisation des systèmes biométriques nous a permis de constaté que la technique du Fuzzy Vault est la mieux indiquer pour la sécurisation des dispositifs biométriques mobiles, du fait de l'utilisation de clés secrète en plus des données biométriques. Une étude détailler sur le Fuzzy Vault a été effectuer et deux PFE ont été proposés, dont un a abouti sur une plateforme exploitable. Cela nous a permis d'effectuer nos tests et de valider nos propositions.

PUBLICATIONS

- [1] Hachemi Nabil Dellys, Layth Sliman, Karim Benatchba. "On efficiency of fuzzy vault algorithm: a Comparative Study of the Different Schemes used in Chaff-Points Generation and Points-Alignment Stages", *Acceptée le 12 septembre 2017, Numéro du volume pas encore communiqué.*
- [2] Hachemi Nabil Dellys., Layth Sliman., Saliha Artabaz., Karima Benatchba., Mouloud Koudil. "Chaff-Points Generation Using Knapsack problem resolution in Fingerprint Fuzzy Vault", 16th International Conference on Hybrid Intelligent System (HIS), Marrakech, Morocco, Nov 2016.
- [3] Hachemi Nabil Dellys, Layth Sliman, Saliha Artabaz, Karima Benatchba and Mouloud Koudil, "Efficiency of Feature Representation on Points-Alignment in Fingerprint Fuzzy Vault" VIIS 2016, Putrajaya, Malaysia Oct 2016.
- [4] Hachemi Nabil Dellys., Noussaiba Benadjimi., Meriem Romaiassa Boubakeur., Layth Sliman., Fathelalem Ali. "Chaff Point Generation by Squares Method Using Composite Representation in Fingerprint Fuzzy Vault", Published in Journal of Information Assurance & Security; 2016, Vol. 11 Issue 1, p1-10, 10p.
- [5] Hachemi Nabil Dellys., Noussaiba Benadjimi., Meriem Romaiassa Boubakeur., Layth Sliman., Fathelalem Ali. "Fingerprint fuzzy vault chaff point generation by squares method". 7th International Conference of Soft Computing and Pattern Recognition (SoCPar), Fukuoka, p357-362, Nov 2015.
- [6] Hachemi Nabil Dellys., Noussaiba Benadjimi., Meriem Romaiassa Boubakeur., Layth Sliman., Karima Benatchba., Saliha Artabaz., Mouloud Koudil. "A Critical Comparison of Fingerprint Fuzzy Vault Techniques". IVIC 2015, Advances in Visual Informatics Volume 9429 of the series Lecture Notes in Computer Science pp 178-188, Nov 2015.
- [7] Dellys Hachemi Nabil, Karima Benatchba, Mouloud Koudil, Ahmed Bouridane. "Threats models on biometric systems: A comparative study", Published in Journal of Information Assurance & Security . 2013, Vol. 8 Issue 2, p137-146. 10p.
- [8] Dellys Hachemi Nabil, Karima Benatchba, Mouloud Koudil, Ahmed Bouridane. "Threats models on biometric systems: A comparative study". Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference, Sao Carlos, p186-191, Nov 2012.

Un système de recherche d'information adaptatif dans un contexte d'intelligence économique

Lilia ZAOUACHE

Directeurs de thèse : Pr. Rachid CHALAL

Équipe : MSI

Résumé— Les entrepôts de données sont nés au sein des entreprises pour répondre à un besoin crucial d'analyse pour l'aide à la décision stratégique. En raison de l'énorme volume de données historiques stockées dans ces entrepôts, les applications OLAP peuvent renvoyer une grande quantité d'informations non pertinentes. Notre objectif est de proposer un système intelligent adaptatif, permettant d'aider les décideurs à trouver les informations pertinentes par rapport à leur profil. Ce papier présente, le contexte et la problématique de la thèse, ensuite, les travaux accomplis durant les années précédentes, et enfin, les perspectives à développer l'année en cours.

I. INTRODUCTION

L'aide à la décision stratégique est le service principal attendu de l'informatique, dans un contexte mondialisé et évolutif. Elle s'appuie principalement sur les informations provenant de l'environnement externe de l'entreprise, en les mettant en perspective avec les informations internes agrégées à un haut niveau.

Pour faciliter et soutenir le processus d'aide à la prise de décision stratégique, les données à analyser sont centralisées de manière uniforme et structurée dans un entrepôt de données [1]. Ces données sont souvent décrites au moyen d'un modèle multidimensionnel [2], basé sur des sujets d'analyses (appelés faits) et des axes d'analyses (appelés dimensions). [3]

Au sein de l'entrepôt, des analyses interactives des données sont effectuées par les utilisateurs via un processus d'analyse en ligne (OLAP On-Line Analytical Processing) [2] [4]

II. PROBLÉMATIQUE

De nos jours, l'analyse OLAP est devenue une tâche difficile pour les utilisateurs, en raison de l'énorme volume de données stockées dans les entrepôts de données ; les applications OLAP peuvent renvoyer une grande quantité d'informations non pertinentes qui pourraient rendre le processus d'exploration de données pas efficace et tardif [5] [6]

De plus, les utilisateurs peuvent avoir besoin de contexte d'analyses spécifiques répondant à des besoins particuliers, voire individuels. [7] [8]

De ce fait, plusieurs approches de personnalisation des systèmes OLAP : approches d'adaptation [6] [8] [9] et approches de recommandation [10] [11], ont émergé au cours de ces dernières années pour aider les décideurs à explorer le grand volume de données multidimensionnelles.

À notre connaissance, il n'existe pas un système de personnalisation OLAP permettant, de fournir au décideur, un contenu adapté à son profil et aussi l'aider dans sa prochaine analyse en lui recommandant des requêtes OLAP selon son profil.

Notre travail de thèse consiste de proposer et évaluer un cadre de personnalisation et recommandation des requêtes OLAP dans les systèmes de gestion de bases de données multidimensionnelles en exploitant le profil utilisateur,

Les points principaux de la thématique abordés sont :

- Définition du profil utilisateur en se basant sur ses contraintes et son contexte actuel au moment des analyses OLAP ;
- Intégration de ce profil en recherchant dans le fichier log, le contenu et les requêtes les plus pertinentes
- Évaluation de l'approche proposée sur une collection de données appropriée.

III. TRAVAUX RÉALISÉS

La première année a été consacrée pour la découverte du domaine de la personnalisation qui est nouveau pour nous, Pour ce faire, nous avons effectué un état de l'art comportant les travaux suivants:

- La personnalisation de l'information
- Processus de la personnalisation : Démarches, techniques et approches[3] ;
- Domaines d'application de la personnalisation;
- Méta modèle du domaine de la personnalisation;

Durant la deuxième année, nous avons effectué :

- Une étude et une synthèse sur les travaux réalisés dans le cadre de la personnalisation des systèmes dans différents domaines tel que : e-learning, e-commerce, RI et les entrepôts de données.

Durant la troisième année,

- J'ai effectué un stage de longue durée (7 mois) dans le cadre de la bourse PNE au sein de laboratoire IRIT à l'université Paul Sabatier de Toulouse, il m'a permis d'affiner et de rediriger mon travail de recherche qui était vaste et confus au départ.

- Une étude approfondie sur le domaine de bases de données multidimensionnelles et la personnalisation des systèmes OLAP.

IV. PERSPECTIVES

Pour l'année prochaine, nous envisageons d'effectuer les travaux suivants :

- Finaliser la rédaction de l'état de l'art portant sur tous le domaine d'étude.
- Rédaction d'un article décrivant le nouveau système permettant de fournir les informations personnalisées dans le contexte d'aide à la décision stratégique.

À long terme, les travaux qui nous restent sont:

- Évaluation de l'approche proposée.
- Rédaction et soumission de la publication.
- La rédaction du manuscrit de thèse.

V. CONCLUSION

À travers ce rapport, nous avons présenté l'état d'avancement de nos travaux de thèse. Après un rappel du contexte et de la problématique traitée, nous avons présenté brièvement les travaux réalisés ainsi que les perspectives fixées pour les années qui suivent.

References

- [1] R. Kimball, M. Ross, "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" (3rd ed.). John Wiley & Sons, Inc. 2013.
- [2] S. Chaudhuri, U. Dayal. "An overview of data warehousing and OLAP technology". ACM Sigmod record, vol. 26, no 1, p. 65-74. 1997.
- [3] M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis. "Fundamentals of Data Warehouses". Springer, Heidelberg. 2000.
- [4] G. Colliat. "OLAP, relational, and multidimensional databases systems". ACM Sigmod Record, vol.25, no 3, p. 64-69, 1996.
- [5] S. Aissi, M.S. Gouider. "Towards the next generation of data warehouse personalization system: a survey and a comparative study". arXiv preprint arXiv: 1208.0203, 2012.
- [6] L. Bellatreche, A. Giacometti, P. Marcel, H. Mouloudi, and D. Laurent, *A personalization framework for OLAP queries*, International Workshop on Data Warehousing and OLAP (DOLAP'05), 2005, pp. 9-18.
- [7] C. Favre, F. Bentayeb, and O. Boussaid, *Evolution et personnalisation des analyses dans les entrepôts de données : une approche orientée utilisateur*, XXVème congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 07), Perros-Guirec, May 2007, pp. 308 - 323
- [8] C. Favre, F. Bentayeb, O. Boussaid, *A User-driven Data Warehouse Evolution Approach for Concurrent Personalized Analysis Needs*, Integrated Computer-Aided Engineering (ICAE), vol. 15, n 1, 2008, pp. 21-36
- [9] I.Garrigós, J. Pardillo, J-N. Mazón, and J. Trujillo, *A Conceptual Modeling Approach for OLAP Personalization*. Conceptual Modeling, ER 2009, LNCS, vol. 5829, Springer, Heidelberg, 2009, pp. 401-414.
- [10] H. Jerbi, G. Pujolle, F. Ravat, O. Teste, *Recommandation de requêtes dans les bases de données multidimensionnelles annotées*. Ingénierie des Systèmes d'Information Vol. 16, N.1, 2011, pp113-138
- [11] A. Giacometti, P. Marcel, and E. Negre, *A Framework for Recommending OLAP Queries*, 11th international Workshop on Data Warehousing and OLAP (DOLAP'08), 2008, pp. 73-80.

Simulation de la valeur des BMO (Business Model Ontologies) en entreprise par les SMA

Hichem_Benaissa_Anouar Badsì

Directeurs de thèse : GHOMARI Abdessamed Réda, Professeur, LMCS, ESI

Équipe : MSI (Management des Systèmes d'Information)

Résumé—Les chercheurs désirent comprendre comment et dans quelle mesure l'application des Technologies de l'Information (particulièrement les Ontologies du Business Model) permet d'améliorer les performances organisationnelles des entreprises. Notre thèse ambitionne de simuler l'impact des BMO à partir d'indicateurs proposés.

L'année en cours a consisté à préparer le développement de cette simulation et à implémenter le modèle de simulation exprimé en BPMN (Business Process Model and Notation).

Mots-clés: Ontologie, Business model, SMA, Simulation, BPMN

I. INTRODUCTION

Les chercheurs ont adopté plusieurs approches pour cerner tous les aspects de l'entreprise [1]. Parmi ces approches on retrouve le Modèle d'affaire ou Business Model. Ce dernier est un modèle abstrait, que les ontologies peuvent aider à représenter.

On retrouve donc des Ontologies du Business Model [2-4] qui fournissent des descriptions abstraites des entreprises dans leur contexte d'affaires, en se concentrant sur ce qui est nécessaire pour créer et transférer de la valeur.

Notre précédent travail, réalisé au cours du mémoire de magister [5] a été l'occasion d'explorer différents cas d'application des BMO et d'établir une liste non exhaustive d'indicateurs pour mesurer l'impact concret des ontologies du business model dans un contexte d'entreprise. Ces indicateurs ont pour but de fournir un cadre pour faciliter l'évaluation de ce type d'ontologies lorsqu'elles sont utilisées dans les modèles d'affaires du monde réel.

Les précédents indicateurs ont été affinés pour faire l'objet d'une publication [6] et former une base solide sur la quelle nous avons poursuivi nos travaux de recherche dans le cadre de la présente thèse.

II. PROBLEMATIQUE

Ces indicateurs étant inhérents, notamment, au degré de consensus concernant la représentation des indicateurs de performance de l'entreprise, au niveau de la standardisation et de la formalisation des processus d'affaires ainsi qu'au niveau de la standardisation des qualifications et des résultats. Cette proposition doit être validée à travers la mesure de ces indicateurs avant et après l'utilisation de l'ontologie du Business Model (BMO) dans une entreprise donnée.

Au vu de ces contraintes de temps et de moyens, nous avons opté naturellement pour l'utilisation d'une approche de simulation pour mener à terme l'évaluation des BMOs dans le milieu des entreprises.

La réalisation d'une telle simulation implique des préparatifs rigoureux, afin d'obtenir des résultats validables, que cela soit pour la définition des indicateurs ou d'un modèle de simulation à implémenter, ou encore, le choix de la technique et des cas d'étude, ou encore, pour, la collecte des informations nécessaires au projet.

III. TRAVAUX RÉALISÉS

A. Première Année

La première étape dans ce travail de thèse fut d'établir les bases théoriques nécessaires pour la recherche, et à collecter les informations nécessaires sur les 3 domaines utiles à notre recherche (Business Model Ontologies, Systèmes Multi-Agents, Key Performance Indicators). Pour cela :

- Un état de l'art a été réalisé sur les ontologies du business model ainsi que sur les systèmes multi-agents.
- Une approche a été proposée pour l'extraction d'indicateurs d'entreprise sensibles à l'utilisation de BMOs.
- Une validation de la pertinence du thème de recherche et des approches proposés a été faite par la rédaction de deux articles (une dans une revue internationale et une dans une conférence internationale):
 - Le premier sur l'approche d'extraction d'indicateurs et la proposition de ces derniers [6].
 - Le second sur la sélection de la méthode de simulation nécessaire à la validation des indicateurs proposés [7].

B. Deuxième Année

Nous avons entamé une phase d'analyse essentielle avant la modélisation et l'implémentation informatique de la simulation. Une première décision lors de cette phase fut de restreindre notre recherche sur la perspective client de l'entreprise, c'est-à-dire l'impact des BMOs sur la gestion de la relation client (CRM). Une recherche bibliographique sur le domaine nous montre l'existence de plusieurs visions différentes de la CRM. Les étapes de réalisation de la simulation ont été redéfinies de manière plus stricte, car la validation du modèle est cruciale avant d'envisager une quelconque implémentation.

Des critères minimaux et idéaux ont été posés afin d'identifier des entreprises potentiellement intéressantes. A titre d'exemple l'existence d'une stratégie CRM et d'un fichier client numérique sont des prérequis obligatoire alors que l'utilisation de canaux de communication multiple et de plusieurs segments de marché ciblés par l'entreprise sont très fortement recommandés.

Un stage de courte durée a été effectué au laboratoire LIRIS de l'Université Claude Bernard Lyon1 sur invitation du

Pr Hassas pour aborder l'aspect des SMA qui était encore non-familier et bénéficier d'avis expert sur le travail réalisé et à venir.

Trois aspects essentiels ont été abordés et éclaircis en cours d'année, à savoir :

- La définition des étapes de réalisations de la simulation et des prérequis au développement.
- La comparaison des méthodologies.
- La comparaison et la sélection d'une plateforme pour l'implémentation de la simulation.

C. Troisième Année

Au cours de la troisième année, nous avons exploré différentes approches pour la conception d'un modèle multi-agent viable.

Notre choix c'est porté sur l'approche théorique, car la littérature du domaine est suffisamment abondante, bien plus accessible que les données d'entreprise et moins spécifique à des cas particulier. Les modèles et descriptions réalisées par les chercheurs sont consensuelles et plus facilement exploitables que des données brutes.

Le modèle réalisé est exprimé en langage standard BPMN pour le rendre compréhensible par des experts d'autres domaines.

Les principales tâches réalisées au cours de cette année sont :

- Le choix des approches de conception et de validation.
- L'étude et la conception d'un modèle de simulation détaillé et plausible.
- L'expression de ce modèle conformément au standard BPMN pour le soumettre à une validation par les experts du domaine.
- La rédaction et la soumission du modèle à un journal [8]

D. Quatrième Année

Cette quatrième année correspondent à la phase d'implémentation du simulateur nommé BMOSIM. La plateforme MaDKit [9] fut retenue comme base pour le simulateur. Un stage d'un mois a été effectué du 19 Mai au 23 Juin 2017 au laboratoire LIRMM de l'université de Montpellier sur invitation du Dr Michel nous a permis de mieux utiliser cette plateforme.

L'architecture du simulateur se compose de trois couches :

- 1- Une couche Interface Utilisateur, pour l'entrée des paramètres de simulation et la récupération des résultats.
- 2- Une couche Agents, comprenant un agent Launcher ayant pour tâche de générer :
 - Les cinq agents du modèle
 - Un pseudo agent faisant office d'interface avec la base de données
 - Un agent Scheduler pour synchroniser la simulation
 - Un agent Watcher pour l'observation et la synthèse des résultats.
- 3- Une couche Données est Connaissances constituée de la base de données, de l'ontologie ainsi que des bibliothèques Java correspondantes.

Les travaux réalisés cette année se récapitulent comme suit :

- La conception d'une architecture pour le simulateur
- La réalisation d'un prototype de BMOSIM
- L'exécution d'un jeu de données de test (Paramètres de simulation, paramètres d'agents, Offres, Business Model)
- La rédaction d'un article sur les premiers résultats et soumission à une conférence internationale spécialisée (fin juillet 2017).
- Rédaction du chapitre « Validation » de la thèse de Doctorat

IV. PERSPECTIVES

La rédaction de la thèse a été entamée pour un dépôt prévisionnel en décembre 2017. Cette dernière étape, inclue la réalisation de tests supplémentaires avec différentes utilisations de la BMO afin d'enrichir le chapitre « implémentation » avec des résultats plus détaillés.

V. CONCLUSION

Au cours de cette année, une première version de BMOSIM a été réalisée puis corrigée au cours d'un stage au laboratoire LIRMM. Un protocole de vente standard a été implémenté pour tester le moteur de simulation. Un article exposant la conception du simulateur et les résultats des premières expérimentations a été soumis à une conférence spécialisée fin Juillet.

References

- [1] J. L. Dietz, "System Ontology and its role in Software Development," EMOI-INTEROP, vol. 5, 2005.
- [2] G. L. Geerts and W. E. McCarthy, "Use of An accounting object infrastructure for knowledge-based enterprise models," AAAI Technical Report, 1999.
- [3] J. Gordijn and H. Akkermans, "Designing and evaluating e-business models," IEEE intelligent Systems, pp. 11-17, 2001.
- [4] A. Osterwalder, "The business model ontology: A proposition in a design science approach," Institut d'Informatique et Organisation. Lausanne, Switzerland, University of Lausanne, Ecole des Hautes Etudes Commerciales HEC, vol. 173, 2004.
- [5] H. B. A. Badsı, "Ontologies & Business Values (Ontologies et Valeur du Business)," Magister Magister, ESI, National Superior School of Computer Science, Algiers, Algeria, 2013.
- [6] H. B. A. Badsı, L. Zemmouchi-Ghomari, and A. R. Ghomari, "How to measure business value of business model ontologies? A proposal," IJBIS (International Journal of Business Information Systems, Inderscience), vol. 19, pp. 241-264, 2015.
- [7] H. B. A. Badsı, A. R. Ghomari, and L. Zemmouchi-Ghomari, "Simuler pour Evaluer les BMOs: Quelle Méthode Choisir ?," presented at the 2nd International Conference on New Technologies & Communication (ICNTC'15), Hassiba Benbouali University of Chlef, Algeria, 2015.
- [8] H. B. A. Badsı, A. R. Ghomari, and L. Zemmouchi-Ghomari, "A CRM Process Model for Agent-Based Simulation" International Journal of Service Science, Management, Engineering, and Technology (IJSSMET), vol. 8(4), pp. 56-82, 2017.
- [9] O. Gutknecht and J. Ferber, "The madkit agent platform architecture," in Infrastructure for Agents, Multi-Agent Systems, and Scalable Multi-Agent Systems, ed: Springer, pp. 48-55, 2001

Framework d'analyse de Big Data basé sur la théorie des ensembles flous

Abid BOUALEM

Directeurs de thèse : Pr. Rachid CHALAL, Pr WK. HIDOUCI
Laboratoire LMCS, ESI (Ecole Nationale Supérieure d'Informatique), 16309, Alger, Algérie.)

Résumé—

Ce rapport décrit l'état d'avancement de la thèse de doctorat initiée en 2014 sous le thème : Framework d'analyse de Big Data basé sur la théorie des ensembles flous.

I. INTRODUCTION

L'accroissement continu des données numériques dans les entreprises et les organismes publics a conduit à l'émergence du «Big Data». Ce concept recouvre les questions de stockage et de conservation de vastes quantités de données, mais aussi celles liées au gisement potentiel de valeur que représentent ces masses de données.

II. PROBLEMATIQUE

Devant l'accroissement massif des volumes de données à traiter et les besoins des entreprises. Les données nécessitent d'être exploitées et analysées en exploitant le vécu (l'historique) afin d'extraire les connaissances et prédire le futur.

Face à ce grand volume de données, l'utilisation des techniques de fouille de données classiques ne permet pas d'extraire le maximum de connaissances à partir de ces données.

- Motivation de relever le défi jusqu'à présent non totalement solutionné d'extraire des connaissances exploitables par les experts non informaticiens,

La théorie des sous-ensembles flous permet d'améliorer les systèmes d'information, notamment pour l'interprétation qualitative et quantitative (numérique). Les informations présentées à l'utilisateur (sous formes de règles par exemple) sont plus compréhensibles et se généralisent mieux.

- Motivation d'utilisation des techniques de logique floue pour l'analyse des données (détection des corrélations cachées, données manquantes).

Dans de nombreux domaines applicatifs, tels que la biologie, la santé, les télécommunications, la vidéo-surveillance, l'énergie, l'environnement, etc., des données sont enregistrées de manière continue. Dans ces cas, le besoin de traitement en temps réel des informations (streaming), représente un défi à explorer

- Motivation des nouvelles technologies Big Data (MapReduce, Hadoop, sqoop, Hive).

Dans ce travail, nous proposons de développer une approche basée sur la logique floue pour l'analyse des données dans un environnement Big Data. L'approche proposée va exploiter la force de traitement des techniques Big Data

comme MapReduce et la puissance d'analyse et d'interprétation de la logique floue (en présence des valeurs vides et imprécises).

Dans le cadre de cette thèse, et dans un premier temps, nous aborderons les problèmes d'analyse dans le Big data, les techniques du data mining.

En particulier, Ce travail portera sur :

1. La conception d'algorithme de data mining adaptative dans les Big Data en utilisant la théorie des ensembles flous ainsi que leur implémentation ;
2. La proposition d'un Framework d'analyse adapté au Big Data en prenant en compte le concept de MapReduce.
3. L'étude de passage à l'échelle du modèle proposé et évaluation des performances par une étude de cas

La validation des algorithmes est prévue pour les cas de gestion de la relation client (CRM) pour aider les entreprises à : trouver, gagner et garder leurs clients. On pourra aussi aborder d'autres utilisations :

1. Churn (la perte de clientèle),
2. Détection des fraudes en temps réel,
3. Détection des attaques des systèmes, base de données d'une entreprise (analyse des logs des applications des serveurs,...)

III. TRAVAUX REALISES

Cette deuxième année de thèse est une suite de la recherche bibliographique et la définition de l'état de l'art.

Un ensemble d'articles [2], [3], [4],[5] pertinents en relation avec le contexte étudié et analysé afin de cerner tous les concepts clés tel que :

- Les trois (3) V du Big Data (Volume, variété, Vitesse),
- La complexité des données manipulées dans les entreprises (plusieurs formes de données selon la nature et la source (application, logs, web))

Une étude approfondie des différents challenges (défi) est réalisée afin de choisir une ou plusieurs pour les futures travaux de recherche. On trouve :

1. **L'hétérogénéité** : afin de produire une quantité importante d'information, l'hétérogénéité est fortement tolérable. Toutefois, les algorithmes d'analyse requièrent des données homogènes. Par conséquent, les données doivent être structurées,

transformées, nettoyées et corrigées avant d'être analysées. Certaines incomplétudes et erreurs sont toujours susceptibles de rester dans les données et doivent être gérées au cours de l'analyse et cela reste un défi important pour les Big Data [1].

2. **Evolutivité** : est le défi majeur pour le Big Data. Pour toute entreprise dans un environnement concurrentiel, elle doit être en mesure de passer à l'échelle très rapidement et de manière flexible, d'où la nécessité d'une solution fiable, et évolutive en terme de stockage et de traitement de grande quantité de données [1].
3. **Performance** : avec internet les données sont générées avec une très grande vitesse. Cela a créé le besoin d'analyse en temps réel ou presque réel. Les traitements effectués par les SGBD mettent un frein pour les solutions d'analyse classiques concentrées autour d'un entrepôt de données [1].
4. **Flexibilité de conception** : les données du Big Data sont sous différentes formes (structurée, semi-structurée et non structurée). La modélisation avec un schéma relationnel fermé est impossible ; au contraire, les entreprises ont besoin d'une conception plus souple et flexible pour que les technologies déployées à leur niveau s'adaptent aux données et non pas l'inverse. Les solutions de big data doivent être en mesure de produire plus avec les données : faire des transactions en temps réel, faire des analyses pour trouver toutes informations/rerelations cachées [1].

Après avoir étudié quelques solutions existantes du Big data, on trouve qu'une maturité est existante dans les couches d'accès aux données et aux stockages par l'existence des API et connecteurs par des constructeurs tel que google, yahoo, Amazon et cloudera et la communauté du projet open source apache (HADOOP). De même pour la couche présentation on trouve les outils BI commercial ou non tel que Tableau pour la visualisation et présentation des données et Talend pour l'intégration.

On trouve plus intéressant d'adresser la recherche dans la modélisation et la proposition de modèle permet au entreprise de mettre en place des plateformes de prise de décision afin :

1. Augmenter les revenus,
2. Amélioration des services proposés en temps réels,
3. minimisation du churn

Et cela par :

L'adaptation des algorithmes traitant des volumes important de données issues des différentes sources (data consommées par les abonnés des opérateurs télécom, localisation des abonnés),

Intégration des données en temps réel tel (rechargement des forfaits pour les abonnés des opérateurs télécom, les positions des abonnés),

IV. PERSPECTIVES

Un travail de conception et d'implémentation sera mené, pour réaliser les couches du modèle présenté, durant l'année en cours de thèse en se basant sur les outils open source ainsi que l'étude d'un cas d'intégration dans une entreprise de télécommunication. Après finalisation du modèle, La soumission d'un article est prévue pour l'année en cours.

Références

- [1] A. A. Divyakant, B. Philip, B. Elisa, D. Susan, D. Umeshwas, "Challenges and Opportunities with Big Data 2011-1" (2011). Cyber Center Technical Reports. Paper 1
- [2] K. Ishwarappa and J. Anuradha, "A Brief introduction on Big Data 5 Vs characteristics and hadoop technology", ICCDC-2015, pp.319-324, 2015
- [3] M. Chen, S. Mao, and Y. Liu, "Big Data : A Survey", MONET 19(2), pp. 171-209, 2014.
- [4] N. Elgendy and A. Elragal, "Big Data Analytics: A Literature review paper", ICDM 2014, LNAI 8557, pp. 214-227, 2014.
- [5] X Wu, X Zhu, G. Wu, and W. Ding, "Data Mining with Big Data" IEEE Transactions on Knowledge and Data Engineering, VOL. 26, NO. 1, 01/2014.

Approches de combinaison d'informations biométriques de l'empreinte palmaire et digitale

Saliha Artabaz

Mouloud Koudil et Karima Benatchba

Équipe : Aucune.

Résumé—Les systèmes biométriques s'appuient sur les caractéristiques de l'être humain pour identifier les individus. Ces systèmes biométriques se basent sur différents types de modalités : morphologiques (empreinte digitale, visage, ...), comportementales (voix, dynamique de frappe, ...) et biologiques (ADN, sang, ...). L'étude des modalités utilisées dans un système biométrique, porte sur l'évaluation des caractéristiques suivantes : l'universalité, l'unicité, la permanence, la mesurabilité, le consentement des utilisateurs et le respect de l'intimité. De plus, le système doit satisfaire les caractéristiques suivantes : performance, protection contre le contournement et extensibilité du nombre d'utilisateurs. Cependant, les systèmes biométriques basés sur une seule modalité sont confrontés à certaines limitations. Cela est dû, généralement, à la non-satisfaction des caractéristiques comme l'universalité, la permanence etc. d'une part et la faiblesse du système biométrique en termes de performances ou d'exposition aux fraudes d'autre part. La multi-biométrie est une solution pour remédier à ces imperfections des systèmes biométriques traditionnels. Elle consiste en la diversification des données biométriques ou les modules de traitement. La multi-modalité est l'une des formes de la multi-biométrie qui consiste en l'utilisation conjointe de deux modalités ou plus. Elle présente à la fois l'avantage de tirer profit des avantages de chaque modalité, mais aussi permet de rendre la tâche plus complexe pour les imposteurs.

I. INTRODUCTION

L'authentification est un service important pour la sécurisation des ressources logiques ou physiques. Les utilisateurs doivent s'authentifier pour exploiter les ressources mises à leur disposition. Les premières méthodes d'authentification reposaient sur les connaissances (mot de passe, informations personnelles) de l'utilisateur. Cette façon de faire peut être facilement compromise si ces informations sont divulguées à une personne étrangère. Une méthode de sécurisation plus rigoureuse repose sur les caractéristiques biométriques des utilisateurs. Ces caractéristiques présentent deux avantages. Le premier est qu'elles sont rarement exposées au vol ou à la perte. Le second est qu'elles donnent avec précision l'identité de la personne ayant utilisé les ressources. Ce qui assure un autre service de sécurité qui est la non-répudiation.

L'étude des modalités utilisées dans un système biométrique, porte sur l'évaluation de leurs caractéristiques. L'automatisation du processus de reconnaissance, sur lequel se base un système biométrique, est un domaine de recherche qui est en pleine expansion. Les chercheurs visent à mettre à disposition un système automatique qui permet de satisfaire les caractéristiques suivantes : performance, protection contre le contournement et extensibilité du nombre d'utilisateurs. De plus, l'intérêt des chercheurs porte sur la proposition de nouvelles méthodes dans les différentes étapes du processus de

reconnaissance : acquisition, extraction des informations biométriques 'modèles', enregistrement et appariement des modèles. Cela est dans le but d'améliorer les performances et d'assurer le meilleur compromis temps de réponse et ressources exploitées. Les systèmes biométriques traditionnels connaissent des limites flagrantes dû à la non-satisfaction des caractéristiques d'une modalité biométrique, d'une part, et la faiblesse du système biométrique en termes de performances ou d'exposition aux fraudes d'autre part.

La multi-biométrie est une solution pour remédier aux imperfections des systèmes biométriques traditionnels. Elle consiste en la diversification des données biométriques ou les modules de traitement. Elle permet de mieux répondre aux caractéristiques déjà citées. La multi-modalité est l'une des formes de la multi-biométrie. Elle consiste en l'utilisation conjointe de deux modalités ou plus. Elle présente à la fois l'avantage de tirer profit de chaque modalité, mais aussi elle permet de rendre le système beaucoup plus robuste face aux fraudes. La reconnaissance en multi-modalité repose sur le traitement de l'information biométrique en passant par quatre niveaux (acquisition, extraction, appariement et décision). La combinaison des informations issues des différentes modalités est possible sur les quatre niveaux.

II. PROBLEMATIQUE

La multi-modalité constitue une solution pour remédier aux insuffisances des systèmes uni-modaux. Elle permet de réduire le risque d'erreur en se basant sur plus d'une modalité des individus (intégration de plusieurs systèmes biométriques). Le recours à cette solution nécessite néanmoins plusieurs considérations durant la conception, la réalisation et le test pour la validation de la solution. Ces considérations peuvent se résumer en les points suivants :

- Le nombre de modalités et éventuellement l'architecture globale qui définit la gestion des flux d'informations biométriques issues des modalités choisies. Des discussions autour de l'hybridation des formes en multi-biométrie sont entamées par les chercheurs et constituent une brique essentielle pour l'étude des relations et l'impact de l'utilisation des informations biométriques ou non-biométriques des individus ainsi que la diversification des modules de traitements ;
- Le mode de reconnaissance visé : identification ou vérification. Le choix de l'un des modes ou des deux influe considérablement sur le schéma global et le processus de test établi ;
- Le niveau et la méthodologie de fusion qui peut faire appel aux performances et aux spécificités de chaque système. Le choix du niveau et de la méthodologie sont étroitement liés. Ce choix sera traduit par l'architecture global des systèmes après fusion ;

- Le processus de test :
 - ✓ Les bases de données utilisées pour les tests et le protocole de test élaboré pour l'expérimentation et la validation du système.
 - ✓ L'évaluation peut tenir compte non seulement des performances, exprimées en fonction des taux de reconnaissance ou d'erreurs, du système après fusion, en comparant aux systèmes d'origine, mais des critères définis selon les choix de conception préétablis : coût, extensibilité, la généralisation (universalité) et l'effet de l'utilisation d'utilisateurs virtuels.

Le travail visé en thèse consiste en la proposition d'une solution multimodale basée sur deux modalités biométriques : l'empreinte palmaire et l'empreinte digitale. La contribution portera sur l'étude et la comparaison de différentes méthodologies de fusion et l'application des choix de conception cités auparavant. Le système proposé mettra en œuvre les quatre étapes du processus pour chaque modalité. De nouvelles approches de fusion seront proposées et testées. Elles interviendront principalement sur les niveaux d'extraction, de comparaison et de décision. Les méthodes proposées devront être mises en œuvre, validées et comparées entre elles. Des critères de comparaison devront être définis afin de définir l'applicabilité des méthodes aux modalités choisies et l'effet des différents paramètres.

III. ETAT D'AVANCEMENT ET PERSPECTIVES

Dans notre recherche, nous nous sommes intéressés à l'étude des méthodes de combinaison. Cette étude visait l'exploration des avancées de recherche à travers la lecture et l'analyse des publications et des articles les plus récents pour pouvoir classer les différentes méthodes. Nous avons aussi eu le besoin d'étaler la recherche dans d'autres domaines comme la classification, vu que la fusion est aussi employée dans ce domaine. L'exigence en termes de performances et d'aisance d'usage qu'éprouve la multi-modalité élargie le domaine de recherche et demande la conception de solutions à la fois robustes et conviviales. Les applications, étant diversifiées, mobilisent le recours à des solutions adaptées au problème de sécurité selon le niveau de risques, les menaces et les exigences du client. Les choix à établir sont nombreux entre modalités, méthodes d'extraction et de classification et méthodes de fusion. Ces choix sont importants et ne sont pas indépendants si on vise une solution optimale. Les décisions prises doivent s'effectuer mutuellement, ce qui augmente le degré de liberté du problème pour la proposition d'une solution adéquate et donc la complexité de sa résolution. Ce qui rend le problème NP difficile. Nous avons donc formulé le problème afin de mettre en place un modèle permettant la génération et la recherche de solutions selon des critères préétablis. Cette plateforme permet de retrouver un système biométrique multimodal optimal et de comparer différentes configurations de solutions du problème d'authentification ou d'identification, sur la base des performances visées. Vu la complexité et la taille des solutions, un appel à une méthode bio-inspirées est ressenti. Cela dans le but d'optimiser et d'accroître l'efficacité des solutions produites

Au cours de notre étude, nous avons opté pour la mise en place d'une plateforme pouvant servir pour la comparaison. La

plateforme repose sur un modèle que nous proposons afin d'établir l'évaluation des systèmes biométriques multimodaux. Ce modèle décrit les différentes décisions devant être prises pour la mise en place et le test d'un système multimodal. Ces décisions interdépendantes sont toutes des éléments essentiels pour l'optimisation des performances. Ce qui nous a emmenés à les impliquer dans l'élaboration de la solution biométrique et d'analyser l'impact de chacune sur les performances. Cela dans le but d'établir une comparaison assez complète des méthodes de combinaison citées dans la littérature.

Au cours de l'année écoulée, nous avons mis en place les deux processus biométriques des deux empreintes. Des tests de classification et de fusion concernant les deux niveaux sont en cours. Les résultats obtenus sont intéressants pour la méthode d'extraction choisie. Une étude approfondie est menée afin de montrer la robustesse de cette méthode dans le mode d'identification et de vérification. Cette étude fera l'objet d'une publication prochainement.

IV. CONCLUSION

La reconnaissance par les caractéristiques humaines devient de plus en plus sollicitée pour la sécurisation de systèmes et l'identification d'individus. Les systèmes biométriques uni-modaux ne peuvent satisfaire tous les critères qui caractérisent un système biométrique parfait. La multi-modalité, une des formes de la multi-biométrie, présente une solution pour améliorer les systèmes biométriques. L'utilisation conjointe des modalités nécessite une conception qui prend en compte les modalités, l'architecture (parallèle, séquentiel ou en cascade) et le niveau de fusion.

Notre domaine d'intérêt porte sur les méthodes de combinaison utilisées pour la fusion de plusieurs modalités. Nous nous sommes intéressés durant cette année à l'étude des méthodes de combinaison. Nous avons constaté que chaque décision prise et chaque choix établi dans la conception d'une solution est à lui seul un paramètre pouvant améliorer ou détériorer les performances. Un modèle prenant en compte ces paramètres devient donc une nécessité afin d'optimiser leur sélection. C'est pour cela que nous proposons de mettre en place une plateforme répondant à ces besoins.

PUBLICATIONS

- [1] Saliha Artabaz, Hachemi Nabil Dellys, Layth Sliman, Karima Benatchba and Mouloud Koudil, "Multibiometrics enhancement using quality measurement in score level fusion", Advances in Intelligent Systems and Computing, ISDA 2016, porto
- [2] Saliha Artabaz., Layth Sliman., Karima Benatchba., Hachemi Nabil Dellys., Mouloud Koudil. "Score Level Fusion Scheme in Hybrid Multibiometric System". IVIC 2015, Advances in Visual Informatics Volume 9429 of the series Lecture Notes in Computer Science pp 166-177, Nov 2015.
- [3] Saliha Artabaz., Hachemi Nabil Dellys., Karima Benatchba., Mouloud Koudil., Ahmed Bouridane. "Palmprint recognition system: Case Study with different databases". Journal of Information Assurance & Security . 2015, Vol. 10 Issue 3, p130-138. 9p.
- [4] Saliha Artabaz., Hachemi Nabil Dellys., Karima Benatchba., Mouloud Koudil., Ahmed Bouridane. "Extraction method of Region of Interest from hand palm: Application with contactless and touchable devices". Information Assurance and Security (IAS), 2014 10th International Conference, Okinawa, p77-82, Nov 2014.

Development of heterogeneous mapping approaches on Networks on chip

Nassima Kadri

Directeurs de thèse : Pr.Mouloud Koudil

Équipe :

Abstract—An overview of our thesis progress will be given in this summary. We recall that the objective of our research is to propose a new fault tolerant technique for network on chip (NoC), while achieving a good trade-off between reliability and performance (latency, power consumption, throughput, area, etc.).

I. INTRODUCTION

Sustain Network on Chip (NoC) functionality in presence of faults has become a great challenge for NoC designer [1]; especially, with shrinking the size of transistor that on one hand, promotes a high integration density on a single chip which increases the complexity of the system on chip (SoC), and on the other hand, the system as well as the NoC will become more and more vulnerable at different types of faults. Therefore, several researches are existing in the literature that consider this parameter. In this Summary, the progress of our thesis will be given, two principal contributions have been achieved this year. The first is the improvement and the submission of an article on fault tolerance simulator, the second one is the preparation of survey paper on the domain of our research.

II. PROBLEMATIC

Prevent a fault tolerant approach based on mapping technique to improve the reliability of NoC while considering system requirements (latency, energy consumption, throughput, etc). For achieving this goal, we propose a multi-objective mapping-based fault tolerant technique.

III. WORKS DONE

In this section, we detail the work done in this year

A. The first contribution

The simulation step is an essential step in our work, for improving performance, evaluating and validating our solution. Despite the importance of reliability, most of the existing open source NoC tools and simulators do not tackle this parameter.

Due to this fact, we have developed a new simulation platform that considers NoC reliability in addition to performance and power consumption evaluation. Our platform named Fault Tolerant Network on Chip Simulator (FTNoCSim), is designed to make easier the implementation and the evaluation of new fault tolerance techniques for three layers of the NoC OSI model [2]: data link, network and transport layers. We have written an article to provide an overview of our developed

platform, entitled **FTNoCSIM: A new platform simulation for evaluating NoC's reliability**.

This paper has been submitted to publication in Integration, the VLSI Journal (Elsevier) in **July 2017**. Currently, the paper is in Under Review step.

B. The second contribution

Writing survey paper is an important step in PhD research, it is very useful to well comprehend our research field, to get a recent and different review of the literature on our topic. Several surveys on NoC fault tolerance are existing in the literature, Radetzki et al [3] present a survey on fault models, diagnosis techniques and fault tolerant techniques in NoC. They take in consideration existing fault tolerance techniques which are based on different types of redundancy: spatial redundancy, temporal and information redundancy for each NoC layers, except system layer.

Three categories of fault tolerance techniques are discussed in [4], i) Fault tolerance at data level or transport layer, ii) Fault tolerance at routing level and iii) Fault tolerance at topology level or physical layer. Recently, a review on fault tolerant approaches used to bypass permanent faults in NoC is presented in [5], fault-tolerant topologies, fault tolerant routing including approaches based on flooding algorithms [6] are discussed in this review.

We have noticed that fault tolerance approaches at NoC system layer, were not considered or well discussed in the surveys cited above.

In our second contribution, we have written a survey paper to adress these approaches, while focusing on fault tolerant mapping. A classification with discussion of each techniques category have also been considered. The paper is sent to my thesis supervisor for correction and suggestions.

IV. PERSPECTIVES

We plane to submit our survey paper in indexed journal, and to implement our fault tolerant technique, just after completing and submitting the final version of our survey.

V. CONCLUSION

In this summary, we have presented an overview of works done during last university year of our thesis inscription. The principal works are, the submission of an article in **Integration, the VLSI Journal (Elsevier)**. The second is the redaction of a survey paper on our research field.

REFERENCES

- [1] C. Constantinescu, “Trends and challenges in vlsi circuit reliability,” *IEEE micro*, no. 4, pp. 14–19, 2003.
- [2] L. Benini and G. De Micheli, “Networks on chips: a new soc paradigm,” *Computer*, vol. 35, no. 1, pp. 70–78, 2002.
- [3] M. Radetzki, C. Feng, X. Zhao, and A. Jantsch, “Methods for fault tolerance in networks-on-chip,” *ACM Computing Surveys (CSUR)*, vol. 46, no. 1, p. 8, 2013.
- [4] J. Montanana, D. de Andres, and F. Tirado, “Fault tolerance on nocs,” in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*. IEEE, 2013, pp. 138–143.
- [5] S. Werner, J. Navaridas, and M. Luján, “A survey on design approaches to circumvent permanent faults in networks-on-chip,” *ACM Computing Surveys (CSUR)*, vol. 48, no. 4, p. 59, 2016.
- [6] M. Pirretti, G. M. Link, R. R. Brooks, N. Vijaykrishnan, M. Kandemir, and M. J. Irwin, “Fault tolerant algorithms for network-on-chip interconnect,” in *VLSI, 2004. Proceedings. IEEE Computer society Annual Symposium on*. IEEE, 2004, pp. 46–51.

Indexation et recherche d'images par fusion d'informations multimodales

Mahdia BAKALEM

Directeurs de thèse : Pr. N. BENBLIDIA

Laboratoire : LRDSI

Co-directeurs de thèse : Pr. S. AIT AOUDIA

Équipe : IMAGE, Laboratoire : LMCS

Résumé—La recherche d'image est devenue l'une des technologies majeures, indispensable dans notre vie. Cependant, le problème du fossé sémantique, due à la difficulté d'associer une sémantique à une image, persiste dans la recherche d'image. L'annotation d'image permet d'améliorer la recherche d'image, en assignant des mots clés, qui décrivent la sémantique, à l'image. Notre premier but est d'améliorer la recherche d'image en focalisant sur l'amélioration de la qualité d'annotation. L'intégration du thésaurus dans le processus d'annotation permet de minimiser les problèmes de synonymie et de polysémie visuelle et textuelle. Dans ce travail, nous avons proposé une nouvelle approche d'annotation d'image basant sur le WordNet.

I. INTRODUCTION

L'image existe depuis plus de trente mille ans, constituait un moyen de communication pour l'être humain avant l'invention de l'écriture.

En effet, l'essor des technologies de l'électronique, de l'informatique et de l'Internet dans des domaines tels que la médecine, les images satellitaires et la cartographie... a provoqué l'accroissement d'images numérisées sur le Web. Ainsi la richesse d'image en information visuelle et sémantique pose le problème du fossé sémantique dans la recherche d'images sur le Web. Entre l'aspect visuel d'image et les concepts humains où la requête est textuelle basée sur la sémantique.

Une des stratégies prometteuse est l'annotation des images (Indexation textuelle sémantique) qui permet d'attribuer des mots clés à l'image afin de décrire la sémantique véhiculée par cette image. Elle consiste à désambigüiser et interpréter les images d'une façon qui assure leur exploitation adéquate dans la recherche d'image. Cette annotation permet de combiner et de trouver des liens entre le contenu visuel et sémantique afin de minimiser le fossé sémantique dans la recherche des images.

Plusieurs travaux ont été effectués dans ce sens afin d'améliorer la qualité d'annotation des images, qui dépend fortement de certains paramètres.

Dans ce but, notre premier travail était une étude comparative des annotations d'images dans l'espace textuel et latent en montrant l'influence du choix des caractéristiques visuelles sur la qualité d'annotation.

Notre objectif principal est d'améliorer la qualité d'annotation afin d'améliorer la recherche d'image. En affinant le processus d'annotation d'image par la corrélation

visuelle-textuelle (Annotation), appliquée d'abord aux classes de régions visuellement similaires (blobs) puis à la nouvelle image, à base d'un thésaurus.

Dans le but de trouver l'annotation pertinente, plusieurs approches ont utilisé le WordNet dans leurs processus d'annotation d'image.

Le WordNet, qui est l'un des plus populaires thésaurus, utilise le réseau sémantique pour l'estimation de similarité sémantique, qui est la mesure de confiance permettant de refléter la relation sémantique entre les sens des mots. Un mot peut avoir plusieurs sens qui due à l'ambiguïté. Le vrai sens sémantique d'un mot dépend fortement du contexte.

Ce rapport est organisé comme suit : section 2 décrit la problématique, section 3 présente le processus d'auto-annotation d'image. Les travaux réalisés sont exposés dans la section 4, la section 5 concerne les perspectives et finalement la conclusion est abordée dans la section 6.

II. PROBLEMATIQUE

Le problème majeur qui se pose dans la recherche d'images sur le web est comment rechercher rapidement et efficacement des images pertinentes parmi une grande masse d'images disponibles sur le web?

Le problème de fossé sémantique, entre les préférences, requêtes textuelles basées sur la sémantique, de la plupart des utilisateurs et la richesse d'image en information visuel et sémantique, persiste dans la recherche d'image sur le Web.

L'annotation d'image permet d'améliorer la recherche d'image sur le Web en assignant la sémantique véhiculé par l'image à travers des mots clés. L'approche courante consiste à extraire et fusionner les informations textuelles et visuelles des images.

Cependant, l'auto-annotation pose des problèmes de synonymie et de polysémie visuelle et textuelle. Tel que : deux images peuvent être sémantiquement similaires, même si les mots qui les annotent ou leurs caractéristiques visuelles ne sont pas identiques.

Notre problématique est la suivante : Comment améliorer l'annotation automatique des images en faisant la corrélation entre l'aspect visuel et sémantique?

III. PROCESSUS D'ANNOTATION D'IMAGE

Le processus d'auto-annotation d'image utilisé dans notre projet de recherche contient deux grandes phases [2]:

A. Apprentissage

elle consiste à construire les blobs, qui regroupent les régions visuellement similaires, à partir d'une base d'apprentissage, et annoter les blobs construits.

1. *Construction des blobs* : cette étape se déroule comme suit :
 - a) Segmenter chaque image du corpus en région.
 - b) Extraire les caractéristiques visuelles de chaque région.
 - c) Catégoriser toutes les régions afin d'avoir des classes de régions visuellement similaires (blobs).
2. *Annotation des blobs* : la corrélation entre l'aspect visuel (blobs) et l'aspect sémantique (mots clés) à travers un algorithme.

B. Traitement d'une nouvelle image

une image collectée (depuis le Web par exemple) sera segmentée et ses régions affectées aux blobs définis dans la phase précédente. La nouvelle image sera annotée par les mots clés des blobs auxquels ses régions appartiennent.

IV. TRAVAUX REALISES

Après avoir passé en revue les travaux relatifs à l'annotation automatique des images, ainsi que tous les travaux connexes nécessaires au développement d'un processus d'auto-annotation d'images durant l'étude théorique, nous avons pu communiquer l'article «**A Novel Image Auto- annotation Based on Blobs Annotation**» [2].

Dans la suite, nous nous sommes consacrés à l'étude comparative sur l'annotation des images dans l'espace latent et textuel en montrant l'influence du choix des descripteurs visuels.

Plusieurs travaux ont été effectués dans le but d'améliorer l'annotation des images. La qualité d'annotation dépend de plusieurs paramètres, et fortement par le paramètre du choix des caractéristiques visuelles (descripteurs visuels).

Une image contient beaucoup d'informations visuelles couleur, texture, forme... mais le bon choix de ces paramètres reste un problème.

Notre étude s'est focalisée sur les descripteurs visuels concernant la texture et la couleur. Comme, les images manipulées sont généralistes, nous pensons que ces descripteurs sont appropriés.

Les résultats obtenus ont fait l'objet d'un papier, intitulé «**A Comparative Image Auto-Annotation** » [3], publié dans IEEE ISSPIT' 2013.

Après cela, nous nous intéressés à la corrélation visuelle-textuelle (Annotation) à base d'un thésaurus WordNet.

L'intégration de WordNet dans le processus d'annotation d'image est appliqué dans les deux phases du processus. La première concerne le processus d'annotation de blobs, qui sont des classes de régions visuellement similaires ; la deuxième phase quant à elle concerne le processus d'annotation d'une nouvelle image.

Les résultats obtenus ont abouti à la soumission d'un papier journal, intitulé «**Contribution to Automatic Image Annotation Using WordNet**», dans lequel nous avons exposé l'amélioration apportée par le thésaurus WordNet dans l'annotation d'images.

Après la soumission de l'article durant le mois d'août, nous nous sommes concentrés sur l'élaboration du plan et la rédaction de la thèse.

V. PERSPECTIVES

Tout au long de cette année, nous allons finaliser la rédaction de la thèse.

Nous allons aussi évaluer le modèle d'annotation d'images en considérant d'autres benchmarks.

Par la suite, nous envisageons l'intégration du processus d'annotation d'image dans le processus de recherche d'image. Cette étape est importante afin de montrer l'apport de l'annotation dans la recherche d'images. La phase d'expérimentation va permettre de comparer et valider l'ensemble du processus.

VI. CONCLUSION

L'idée principale de notre travail est d'apporter une contribution dans la recherche en améliorant l'annotation des images. Notre approche s'est focalisée sur l'intégration d'un thésaurus, WordNet, dans le processus d'annotation d'image afin de minimiser le problème du fossé sémantique.

Références

- [1] M. Bakalem; N. Benblidia; S. Oukid. Latent semantic analysis based image auto annotation; Proceeding of the International Conference on Machine and Web Intelligence (ICMWI); IEEE, pages 460-463, Algiers, Algeria. October 03-05, 2010.
- [2] M. Bakalem, N. Benblidia, S. Ait-Aoudia. A Novel Image Autoannotation Based on Blobs Annotation; Proceeding of the Advances in Intelligent and Soft Computing 2011, International Conference on Image Processing and Communications Challenges 3, Springer, pages 113-122, Bydgoszcz, Poland, September 07-09, 2011.
- [3] M. Bakalem, N. Benblidia, S. Ait-Aoudia. A comparative image auto-annotation; Proceedings of The 13 IEEE International Symposium on Signal Processing and Information Technology, IEEE ISSPIT 2013, December 12-15, 2013, Athens, Greece.

Impact des réseaux sociaux d'entreprise dans le partage des informations dans les organisations

Hamza LADEG

Directeurs de thèse : Pr Rachid CHALAL

*LMCS Laboratory, ESI (National School of Computer Science), 16309, Algiers, Algeria
H_ladeg@esi.dz*

Résumé

Les Réseaux Sociaux d'entreprises sont définis par la majorité des auteurs comme des outils technologiques qui permet la collaboration entre collègues, capturer et transférer les connaissances et le savoir-faire, augmenter la créativité et l'innovation par la création des liens ainsi que de la conversation.

Pour notre travail nous avons choisi un parmi ces outils qui existent sur le marché pour faire l'expérimentation sur deux groupes d'individus à fin de réaliser un travail collectif au niveau d'un établissement pour voir comment ce fait le partage des connaissances (tacites ou explicites) selon le modèle SECI de Nonaka [1], ainsi le type des connaissances partagées.

1. Introduction

Les technologies dites de réseaux sociaux d'entreprise sont considérées comme des supports pouvant permettre l'animation des réseaux sociaux d'une organisation, soit pour privilégier la communication interne pour améliorer la collaboration entre collègues [Ganesh 2], créer des liens ainsi que de la conversation afin de capturer les connaissances et le savoir-faire et d'autre part, pour augmenter la créativité et l'innovation, soit en mettant l'accent sur la communication externe et le management de la relation clients [Garnier 3]. Les réseaux sociaux sont également de plus en plus utilisés par les entreprises dans la recherche de nouvelles compétences [Girard 4].

En créant un RSE, les individus et leurs connaissances sont mis en valeur. Ces individus ne sont pas classés selon leur position hiérarchique mais par leurs domaines d'expertise.

Le RSE favorise l'échange entre les individus et leur permet de s'épanouir à travers son identité professionnelle. Pour notre expérimentation, nous nous intéressons donc à voir comment et quelles sont les connaissances partagées à travers le RSE, c.-à-d. est-ce que toutes ces connaissances sont valides ?

2. Problématique

Le RSE est pour l'entreprise un nouvel outil de partage de l'information beaucoup plus dynamique que l'email. Il favorise les échanges entre salariés en permettant aux utilisateurs de poster des informations via leurs profils, de stocker, partager et rendre disponible virtuellement des dossiers, de réaliser un travail d'équipe en temps réel ou différé... Il remplace l'email et évite la multiplication des envois de fichiers lourds et la saturation des serveurs de mails. C'est un outil de travail collaboratif qui permet de partager les connaissances et le savoir-faire entre les individus.

La question qui se pose est quelle est la contribution des Réseaux sociaux d'entreprise dans le partage des connaissances valides.

3. Travaux réalisés :

Le travail réalisé est l'implémentation d'un RSE au niveau d'un établissement (université) pour réaliser un travail collaboratif, qui regroupe plusieurs individus (projet ou travail à réaliser, groupe de travail et logiciel RSE). Pour notre cas nous avons utilisé deux groupes d'individus différents pour le même projet avec quelques tâches supplémentaires pour le deuxième groupe. Chaque groupe contient sept personnes. Dans les deux groupes les individus utilisent les principales fonctionnalités de la plateforme RSE (documents, vidéos, photos, wikis, forum de discussion, etc....) pour partager leurs connaissances, afin de réaliser le travail demandé d'une façon collaborative et dans un délai déterminé.

Résultats :

D'après les différents modes de partage de connaissances (documents, vidéos, photos, wikis, forum de discussion...) utilisés par les utilisateurs du RSE dans les deux travaux pratiques, nous avons constaté quatre types de connaissances partagées à travers la plateforme collaborative pour les deux groupes : connaissances valides complètes, connaissances valides insuffisantes, connaissances valides hors projet et connaissances non valides.

Pour notre travail nous avons analysé les connaissances partagées pour chaque groupe.

Le premier groupe

- 46% connaissances valides
- 23% connaissances valides insuffisantes.
- 18% connaissances valides hors projet de travail.
- 13% connaissances invalides.

Le deuxième groupe

- 43% connaissances valides
- 29% connaissances valides insuffisantes.
- 14% connaissances valides hors projet de travail.
- 14% connaissances invalides.

La participation des individus pour les deux groupes dans le partage des connaissances est :

Le premier groupe :

- 80% qui ont participé dans le partage des connaissances.
- 20% ne participent pas.

Le deuxième groupe :

- 60% qui ont participé dans le partage des connaissances.
- 40% ne participent pas.

Le taux des connaissances partagées dans chaque groupe :

- 75% de connaissances partagées dans le premier groupe.
- 25% de connaissances partagés dans le deuxième groupe.

4. Perspectives

Suite aux deux expériences que nous avons menées, nous avons constatés les éléments suivants :

1- le processus du transfert des connaissances selon le modèle SECI est respecté avec l'utilisation du réseau social d'entreprise pour les quatre modes de transfert, à travers les différentes fonctionnalités du RSE tel que (mur, wiki blogs forums de discussion, documents photos vidéos, ...etc.).

2- l'utilisation du RSE dans le travail collaboratif apporte une amélioration dans la durée de la réalisation du travail lorsque la base de connaissances de la plateforme est remplie. En effet, nous avons relevé que le deuxième groupe a pris moins de temps pour réaliser le travail demandé malgré qu'il y'avait des tâches supplémentaires ajoutées par rapport au premier groupe. Ceci s'explique par le fait que pour :

- Le premier groupe la base de connaissances était vide (problème du démarrage à froid). Donc, pas de connaissances à partagées. Dans ce cas, le travail collaboratif tarde à se mettre en place.
- Pour le deuxième groupe, nous avons constaté moins d'interactions car la base de connaissances de la plateforme est déjà remplie par le premier

groupe. Ce qui a permis au deuxième groupe de prendre moins de temps pour réaliser le travail.

3- le seul problème soulevé lors des deux expériences et que les individus des deux groupes partagent parfois des connaissances invalides. Ce qui met en péril le travail à réaliser par le groupe.

5. Conclusion

A la lumière de cette étude, nous proposons d'introduire dans le RSE une fonctionnalité pour la validation des connaissances avant de les rendre accessibles pour être partagées par les individus formant le groupe de travail. Cette fonctionnalité devra mettre en œuvre des mécanismes de validation des connaissances

Bibliographie

- 1 Nonaka, I. et Takeuchi, H. "The knowledge-creating company : how Japanese companies create the dynamics of innovation", Oxford University Press, New York. 1995
- 2 Ganesh J., Padmanabhuni S., (2007), « Web 2.0: Conceptual Framework and Research Directions », AMCIS Proceedings, Colorado,
- 3 Garnier, A., & Hervier, G. (2011), Le réseau social d'entreprise, Collection Management et Information, Hermès Lavoisier.
- 4 Girard A., Fallery B., Rodhain F., (2011), « L'apparition des médias sociaux dans l'e-GRH : gestion de la marque employeur et e-recrutement », 16ème Congrès de l'Association Information et Management, 25-27 mai, Saint Denis, Ile de La Réunion.

Le routage dans les réseaux sur puce (NoCs)

Benmessaoud Gabis Asma

Directeurs de thèse : Prof. Mouloud Koudil

Équipe : CoDesign

Abstract—Les réseaux sur puce (NoC) est une nouvelle technique de communication dans les systèmes sur puces (SoC). Celle-ci se base sur le principe de routage des données entre une source et une destination. Pour cela, un routeur est attribué à chaque composant du réseau. Etant dans un environnement miniature, alors les algorithmes de routage doivent répondre à certaines contraintes spécifiques. En effet, dans un NoC un protocole de routage ne doit pas être gourmand en consommation d'énergie et en consommation d'espace. De plus, il doit être capable d'éviter les deadlock, les livelock, les zones de congestion et qu'il soit tolérant aux fautes afin d'assurer des temps de latence raisonnables et un taux de données transmises maximum. Notre problématique étant situées dans ce contexte, nous expliquons alors dans ce rapport quelles sont les étapes suivies et les travaux réalisés jusque-là afin de répondre à l'objectif de notre thèse.

I. INTRODUCTION

Les SoCs sont une technologie récente qui rentre sur le marché de l'électronique pour faciliter plusieurs traitements. Elle consiste en l'intégration d'un système complet sur une seule et unique puce. Ce système peut comprendre divers composants (appelés IP : Intellectual Property) : processeurs, mémoires, réseaux d'interconnexions, circuits spécifiques É etc. La communication entre ces différents composants se faisait autrefois via des bus de connexion sous différentes topologies. Par la suite, et afin d'accélérer le temps de traitement un nouveau concept a rejoint le monde des SoC à savoir : Les NoCs (ou Network on Chip).

Ces derniers sont encore au stade de la recherche, qui elle, touche à plusieurs domaines : Le mapping, le routage, la fiabilité ainsi que l'ordonnancement.

Notre thèse se focalise sur l'aspect *Routing* dans laquelle on aura à étudier les techniques de routage dans les NoCs ainsi que les objectifs qu'ils tentent d'atteindre.

En effet, L'exécution d'une application au niveau des puces se fait sur plusieurs composants qui communiquent entre eux via des protocoles de routage.

De ce fait, le routage constitue une partie importante dans l'architecture des réseaux sur puce, où il permet de procéder à la recherche d'une route entre une source (composant émetteur) et une destination (composant récepteur) de la manière la plus optimale possible. Son but principal est d'essayer de trouver le chemin qui permet de maximiser les performances du réseau tout en minimisant la quantité d'énergie consommée et d'éviter un maximum d'obstacles possibles.

Notre travail se focalise sur l'aspect *Routing* sur lequel nous avons avancé en suivant le plan ci-dessous :

- Etude du routage dans l'environnement des NoCs,
- Faire sortir ses différents objectifs et établir une classification : approches mono objectif Vs. Approches multi-objectifs

- Proposition une solution de routage multi objectifs avec optimisation des performances.

II. PROBLÉMATIQUE

Plusieurs algorithmes de routage ont été développés pour améliorer la transmission de données sur un réseau sur puce. Certains d'entre eux tentent d'améliorer la consommation d'énergie, d'autres visent à améliorer la latence ou encore le taux de données transmises. Pour cela divers techniques sont employées selon que l'algorithme suive une approche déterministe ou une approche adaptative. Cependant, dans chacune des situations, des problèmes ralentissant le routage peuvent subvenir à savoir : la congestion, le deadlock, le livelock et la présence de pannes. Pour les éviter, plusieurs techniques sont employées : utilisation des canaux virtuels, utilisation des tables de routage, etc. Néanmoins, ces solutions sont coûteuses en termes de ressources consommées, et n'arrive à satisfaire qu'un seul objectif à la fois au détriment des autres.

C'est donc dans ce contexte que s'insère l'objectif de notre thèse, où il est demandé de développer une approche de routage multi-objectifs (qui serait capable d'éviter le deadlock, le livelock, les zones congestionnées et tolérante aux fautes) permettant de maximiser la quantité de données transmises tout en assurant des performances raisonnables sur le réseau.

Cette approche doit être basée sur des heuristiques, métaheuristiques et/ou des techniques d'optimisation multi-objectifs afin d'éviter la consommation excessive des ressources tout en garantissant la bonne transmission de données.

III. TRAVAIL RÉALISÉ

Afin d'atteindre l'objectif de la thèse, les travaux suivants ont été réalisés :

A. Classification des protocoles de routage

Un travail d'état de l'art et de classification des protocoles de routage dans les NoCs a été effectué. La classification s'est basée sur 4 critères qui sont :

- Les protocoles Deadlock free;
- Les protocoles Livelock free;
- Les protocoles évitant la congestion;
- Les protocoles tolérants aux fautes.

Dans le cadre de ce travail, nous avons montré les techniques employées par chaque algorithme pour atteindre un ou plusieurs objectif(s), nous avons mis en avant les avantages et les inconvénients de chaque technique et nous avons cité un certain nombre de pistes qui pourraient constituer des solutions pour améliorer l'efficacité du routage sur un NoC.

Ce travail a été publié dans Elsevier - Journal of System Architecture (JSA).

B. Algorithme de routage à base d'heuristiques

Une approche de routage à base d'heuristique a été développée. L'algorithme réalisé est nommé HRA (Heuristic based Routing Algorithm for NoCs).

Le développement de cet algorithme a été effectué sur un simulateur de NoCs haut niveau appelé : Nirgam. L'analyse hardware quant à elle a été réalisée sous Synopsys Design Compiler.

L'algorithme HRA a les propriétés suivantes:

- Inspirée de l'algorithme de recherche dans un graphe A*;
- Basé sur une heuristique : un principe très peu utilisé dans le domaine du routage dans les NoCs;
- N'utilise aucun canal virtuel ce qui permet de réduire le nombre de buffers utilisés;
- Utilise une fonction de coût bi-objective.

Les métriques observées pour mesurer les performances de l'algorithme HRA sont :

- La latence dans le cas d'un réseau sans pannes;
- La fiabilité en situation de pannes dans le réseau;
- La coût matériel en espace et en énergie consommée.

Ce travail a été publié dans IEEE MCSoc 2016 (IEEE 10th International Symposium on Embedded Multicore/Many-core Systems-on-Chip).

Une version étendue de ce travail a été soumise sur la revue "IEEE Transactions on Multi-Scale Computing Systems". Actuellement, nous préparons une réponse à un "Minor revision".

C. Algorithme de routage multi-objectifs

L'autre idée de routage qui a été proposée est basée sur le principe dit "application-specific". A partir d'un graphe de tâches correspondant à une application donnée, une ou un ensemble de solutions sur mesure sont générées. Le calcul de ces solutions est basé sur une fonction multi-objectifs prenant en considération: les pannes des liens, la congestion et le deadlock. Afin de fournir plusieurs solutions possibles, des métaheuristiques sont utilisées. Sur la base de cette idée, nous avons pu développer deux techniques déterministes différentes.

- Un algorithme minimal basé sur les algorithmes génétiques appelé MORGA;
- Un algorithme non-minimal basé sur GRASP ;

De manière similaire que le HRA, le routage de MORGA et de l'approche basée sur GRASP a été testé sous Nirgam où les métriques observées sont: la latence, le débit ainsi que le taux de réception. Afin d'estimer le coût des tables de routage utilisées en termes d'espace occupé, une étude hardware a été effectuée sous Synopsys Design Compiler.

Les tests ont révélé des résultats intéressants en termes de latence et de fiabilité aussi bien en absence ou en présence de pannes. De plus, les solutions Pareto optimales offrent un meilleur débit.

L'idée générale a été présentée lors du congré de la ROADEF

qui s'est tenu à Compiègne en Février 2016.

L'article global de MORGA a été soumis dans la revue Microprocessors and Microsystems d'Elsevier. Nous avons soumis nos réponses à un Major revision le 15 Août 2017.

D. Rapport de thèse

Rédigé et en cours de correction par les directeurs de thèse.

IV. EN RÉSUMÉ ...

- Extension de l'article HRA: Réponse au "Minor revision" en cours;
- MORGA: Attente d'un retour après une réponse à un "Major revision";
- Rapport de thèse: Lecture et correction des directeurs de thèse en cours.

V. CONCLUSION

Dans le présent rapport, nous avons résumé le déroulement de la thèse : sa problématique, le plan suivi ainsi que les travaux réalisés jusque-là. Le rapport de thèse est rédigé et est en cours de correction.

REFERENCES

- [1] A. Benmessaoud Gabis, M. Sevaux, P. Bomel, M. Koudil and K. Benatchba, Heuristic based Routing Algorithm for Network on Chip, IEEE 10th International Symposium on Embedded Multicore/Many-core Systems-on-Chip), Sept 2016.
- [2] A. Benmessaoud Gabis and M. Koudil, NoC routing protocols - Objective based classification, Journal of Systems Architecture, 66:14-32, 2016.
- [3] A. Benmessaoud Gabis, M. Sevaux and A. Cerqueus, Technique d'optimisation multi-objectifs à base de métaheuristiques pour résoudre le problème de routage dans les réseaux sur puces (NoCs), ROADEF: Recherche Opérationnelle et d'Aide à la Décision, Feb 2016, Compiègne, France.
- [4] A. Benmessaoud Gabis and M. Sevaux, Heuristic based routing algorithm for a Network on Chip (NoC), EURO: 27th European Conference on Operational Research, Jul 2015, Glasgow.

Un système de recommandation de contenu basé sur l'analyse du contenu social

Yamina Aissaoui

Directeurs de thèse : Faïçal AZOUAOU (ESI)

Codirectrice de thèse : Lamia Berkani

Résumé— Les systèmes de recommandation jouent un rôle de plus en plus important dans notre vie quotidienne [1]. Nous intéressons aux systèmes de recommandation de contenu basé sur l'analyse de contenu social, dans e-learning. Dans ce chapitre nous essayons d'exploiter les différents aspects de système de recommandation pour trouver une problématique précise a travaillé sur dans le chapitre de contribution.

I. INTRODUCTION

De nos jours, nous pouvons trouver toute chose sur internet. La quantité de contenu est énorme. Cependant, cela pourrait poser un problème car les utilisateurs passent beaucoup de temps pour rechercher et trouver des produits qui correspondent à leurs besoins, des questions majeures sont posées, sur quelle page web me rendre pour trouver l'information que je recherche ? Quels sont les films que je n'ai pas encore vus et qui pourraient me plaire ? Quelle musique découvrir ? Quels items lire pour enrichir ma connaissance dans mon domaine de recherche ? Pour répondre à ces questions les systèmes de recommandation sont apparus.

D'autre part ces dernières années, nous avons vu la croissance énorme des utilisateurs dans les réseaux sociaux en ligne (OSN) comme Facebook, Twitter etc. Ceci devient la raison principale de permettre le web en tant que méthode populaire pour examiner les données. Nous intéressons aux systèmes de recommandation de contenu basé sur l'analyse de contenu social[2].

L'objectif de cette partie de thèse est de présenter les fondements du domaine des systèmes de recommandation. Il est organisé comme suit. Dans le premier chapitre, nous présentons les concepts de base et des définitions. Ensuite, nous donnons dans le chapitre 2 un aperçu des techniques des systèmes de recommandation dans e-learning (technology enhanced learning) et les méthodes les plus connues pour l'évaluation de ces techniques et enfin dans le chapitre 3 nous présentons la modélisation de l'apprenant avec les différentes techniques, et les approches de récupération des données.

II. PROBLÉMATIQUE :

Les algorithmes de recommandation sont devenus un mécanisme nécessaire pour e-learning qui recommandent des items tels que des cours, e-learning recommande généralement des recommandations basées sur un filtrage collaboratif ou hybride qui suggère des items favorisés par des utilisateurs similaires. Les algorithmes de filtrage

collaboratifs représentatifs comprennent la matrice de factorisation à faible niveau. Cependant, la plupart des systèmes de recommandation souffrent du problème d'espace des données (data sparsity), où le nombre d'items consultés par un apprenant (par exemple, rating) est souvent très faible par rapport au nombre total d'items (habituellement des centaines de milliers à des millions ou même des milliards). Le problème d'espace des données (data sparsity) peut affecter de manière significative la performance des méthodes basées sur filtrage collaboratif, telles que la matrice de factorisation à faible niveau, principalement en deux raisons: la surutilisation « over-fitting » où des données insuffisantes ne sont disponibles pour la formation des modèles, et le problème du démarrage à froid "cold start" dans lequel les systèmes de recommandation ne peuvent pas faire de recommandations pour les nouveaux utilisateurs lorsqu'il n'y a pas de données de comportement historiques à collecter[3].

Notre objectif dans cette thèse est de :

1. proposer une amélioration des systèmes de recommandation actuels en exploitant les différentes données sociales existantes (les posts sur les réseaux sociaux, micro blogs, commentaires...) ;
2. Minimiser la recommandation explicite et maximiser la recommandation implicite.
3. Trouver une solution pour le problème de démarrage à froid.
4. Trouver une solution pour le problème d'espace de données

III. TRAVAUX RÉALISÉS :

Le système de recommandation suggère à chaque personne du contenu qui devrait l'intéresser en s'exploitant le contenu web 2.0 (social) .

Pour atteindre notre proposition nécessite de répondre aux questions de recherche suivantes :

- 1- Comment exploiter les données du web 2.0 pour améliorer la qualité de la recommandation
- 2- Comment récupérer et exploiter des données sociales provenant de sources différentes et représentées dans des formats différents.
- 3- Quelle architecture technique adopter pour ce système de recommandation ?
- 4- Quel est le modèle utilisé pour représenter le profil de l'internaute utilisateur de notre système de recommandation

Pour reprendre à ces questions nous avons présenté dans le premier chapitre les fondements de système de recommandation, les définitions, les types et les techniques de recommandation, et les domaines d'application des systèmes de recommandation.

En raison de la croissance rapide dans les ressources internet spécialement dans le domaine d'étude, les systèmes conventionnels e-learning ont plusieurs problèmes alors nous choisissons de travailler sur le domaine de e-learning [4].

D'autre part nous y avons montré les travaux les plus importants dirigés dans chaque technique de recommandation « approche basé sur le filtrage collaboratif, approche basé sur le contenu, approche basé sur l'exploitation des règles d'association, et approche basé sur l'analyse de contenu social » avec les avantages et les limitations de chaque approche.

Avec le développement du réseau social ces dernières années, comment utiliser l'information sur les réseaux sociaux est devenu un sujet brûlant a cette cause nous concentrons sur les approches basé sur l'analyse de contenu social, plusieurs approches sont proposées pour personnaliser les informations sur les réseaux sociaux :

- Des approches basées sur l'influence des amis sociaux.
- Des approches pour la recommandation des top-k item qui intègre l'information sociale.
- Des approches basées sur les tags d'articles puissent fournir des informations importantes.
- Des approches de recommandation basée sur la similitude entre utilisateurs et utilisateurs avec la capacité de confiance.
- Des approches sont développé un cadre de factorisation matricielle avec régularisation sociale.
- Des approches basées sur l'analyse de sentiments.

Puis nous avons présenté un cadre méta-niveau générique simplifié (GML) qui peut être utilisées pour délimiter différentes classes de systèmes de recommandation TEL.

Finalement nous montrons les méthodes fréquemment utilisé pour l'évaluation de ces systèmes « Exactitude des prédictions, Précision du classement » et les techniques de modélisation de l'apprenant plusieurs modèles sont présenté « Le modèle de superposition, model de stéréotypes, modèles d'idées fausses et de connaissances erronées, Techniques d'apprentissage par machine et basée sur l'ontologie » nous choisissons de travaillé avec le modèle basé sur l'ontologie.

Enfin pour reprendre à la question comment collecter les informations nécessaire pour construit ce model ? Trois approches sont présentées : explicite, implicite et hybride. Nous choisissons de travaillé avec l'approche hybride parce qu'il combine les avantages des profils d'utilisateur implicites et explicites[5].

IV. PERSPECTIVES :

Comme perspectives, nous nous sommes particulièrement intéressés aux approches basés sur l'analyse de réseaux social, alors nous proposons une approche basé sur :

- Une double similarité basé sur : niveau d'étude et similarité basé sur l'analyse des sentiments ;
- Avec un modèle de confiance c'est-à-dire un apprenant qui est un expert dans le domaine ou un enseignants spécialisé dans cette domaine ;
- Modification de l'étude par exemple : si une recherche est publié en 2011 est consulté par l'apprenant et en 2015 une mise à jour ou modification de cette recherche est publié le système recommande cette recherche.

V. CONCLUSION :

Nous essayons dans ce chapitre d'étudier les différents aspects de système de recommandation, notre objectif principale dans cette chapitre et de trouver une amélioration des systèmes de recommandation actuels capable de prouve une haute qualité de recommandation en se base sur le contenu de réseau social.

VI. RÉFÉRENCE :

1. Salama, A.A., S.A.E. M. E., and M.M. Lotfy, *Review of Recommender Systems Algorithms Utilized in Social Networks based e-Learning Systems & Neutrosophic System*. Neutrosophic Sets and Systems, 2015. **8**(32).
2. Nikam, Y. and M.B.Vaidya, *Survey On Service Recommendations Techniques*. IJARIII, 2016: p. 464.
3. Chrysafiadi, K. and M.V. Cham, *Student Modeling for Personalized Education: A Review of the Literature*. *Advances in Personalized Web-Based Education*. Springer International Publishing, 2015: p. 1-24.
4. Khribi, M., et al., *Recommendation Systems for Personalized Technology-Enhanced Learning. Ubiquitous Learning Environments and Technologies*. Berlin, Heidelberg, Springer Berlin Heidelberg, 2015: p. 159-180.
5. Klačnja-Milićević and M.I. A., *Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions*. " Artificial Intelligence Review 44, 2015. **4**: p. 571-604.

La sécurité dans les réseaux centrés contenu

Ryma BOUSSAHA

Directeurs de thèse : Yacine CHALLAL, Abdelmadjid BOUABDALLAH

Équipe : SURES

Résumé—L'importance prospère des architectures orientées contenu a convié la communauté de chercheurs à établir une mise à niveau de l'Internet afin de réduire les coûts engendrés par la prolifération du trafic notamment le streaming vidéo. Dans ce projet de thèse, nous étudions le problème de sécurité de cette nouvelle infrastructure de réseau connue sous le nom de réseau centré contenu (ICN, CCN ou NDN). Nous nous sommes penchés principalement sur l'authentification du codage réseaux dans les NDN. Nous avons proposé un programme MIP optimisant le coût engendré par les opérations de codage, de signature et de vérification. Nous avons validé notre proposition par des tests d'évaluation des performances et nous avons montré que notre solution assure un bon compromis entre deux solutions triviales proposées dans [1] et [2].

I. INTRODUCTION

Le codage réseau [3] est une nouvelle technique algébrique qui permet d'améliorer les performances des réseaux en introduisant le calcul au niveau des routeurs. Avec l'application du codage réseau, les données qui transitent sous forme de paquets vont être combinés, codés et décodés à travers tout le réseau. Le fusionnement des techniques de codage réseau avec le principe des réseaux centrés contenu est une rocade que nous avons choisie dans le but de profiter de la redondance des données et des fonctionnalités de mise en cache afin de maximiser l'exploitation de la bande passante dans le réseau et d'assurer une meilleure qualité de service.

En effet, les travaux de recherche sur les CCNs [4] se répartissent en cinq axes différents. Ces différents axes sont illustrés dans la Figure I. Dans notre travail, on s'intéresse plus particulièrement aux modèles de sécurité et aux stratégies de mise en cache. Dans le premier nous nous sommes concentrés sur le problème d'authentification du codage réseau. Et concernant le deuxième axe nous nous sommes concentrés sur la proposition de stratégie de mise en cache des messages codés.

En fait, en dépit de tout ses avantages, le codage réseau soulève un problème de sécurité critique, il est très vulnérable aux attaques de pollution [5]. Si un nœud malicieux réussit à s'introduire dans le réseau et à infiltrer des paquets erronés, alors ces paquets vont être combinés en aval avec d'autres paquets et ils vont rapidement polluer tout le réseau. Un adversaire qui compromet une seule communication peut réussir à falsifier tous les résultats du codage réseau vu sa nature diffuse. Pour remédier à ce problème les nœuds doivent détecter l'altération au premier pas, avant sa propagation dans le réseau, ceci est appelé "*hop-by-hop containment*". Les schémas de signatures homomorphiques [6] ont été proposés pour assurer l'intégrité et l'authentification avec ce type d'attaque. Avec le schéma de signature homomorphique, on peut effectuer des opérations sur les chiffres (signatures) sans divulguer le texte en clair. La signature d'une combinaison linéaire de plusieurs paquets représente en elle-même une combinaison des signatures de

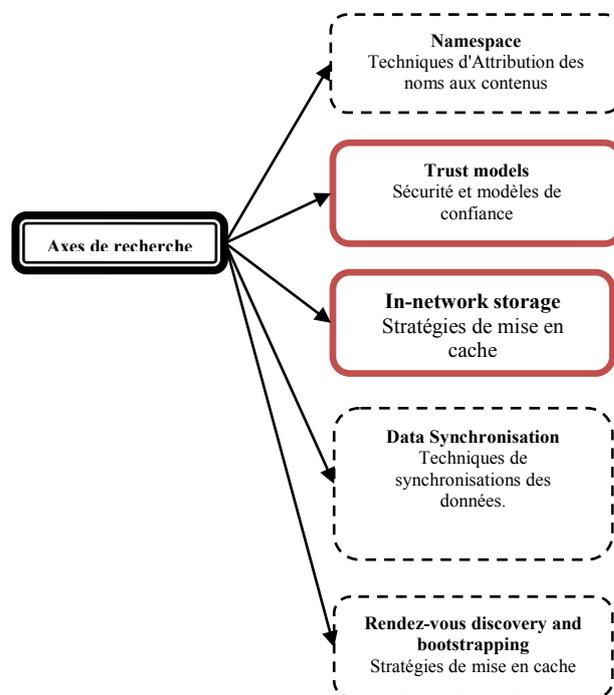


Figure I : Axes de recherche des réseaux centrés contenu.

ces paquets. Cependant, l'utilisation d'un tel schéma de signature avec le codage engendre des coûts de calcul très importants. Le but de notre travail est de proposer une optimisation multi-objective assurant conjointement un coût de calcul minimal et un niveau de sécurité maximal.

D'après l'échéancier de la thèse illustré ci-dessous, je me situe à l'étape T5 qui consiste en l'évaluation des performances des solutions proposées et la comparaison aux solutions de l'état de l'art.

	6 mois	12 mois	18 mois	24 mois	30 mois	36 mois	42 mois	48 mois
T1	■							
T2		■						
T3			■					
T4			■	■	■	■		
T5						■	■	
T6								■

Figure 2: Echéancier prévisionnel de la thèse.

T1: Etude de l'état de l'art sur la sécurité dans les réseaux centrés contenu (6 mois).

T2: Analyse, modélisations et comparaison des solutions existantes.

T3: Apprentissage d'un simulateur de réseaux centrés contenus (6 mois).

T4: Proposition de solutions de sécurité, efficaces et robustes tout en respectant les impératifs des CCNs (24 mois).

T5: Evaluer les performances des solutions proposées (12 mois).

T6: Rédaction de la thèse (6 mois).

II. PROBLEMATIQUE

Le coût de codage des paquets et le coût de calcul des signatures homomorphiques qui permettent d'assurer l'authentification peuvent être contraignant. Donc la question qu'on se pose ici est comment améliorer les performances du réseau en y appliquant le codage réseau tout en assurant un bon niveau de sécurité?

Dans le but de répondre à ce problème nous avons proposé un modèle d'optimisation multi-objective basée sur la programmation linéaire, et plus particulièrement un programme MIP. Ce programme va déterminer l'ensemble des nœuds de confiance qui vont effectuer le codage, la vérification et le calcul des nouvelles signatures homomorphiques. Nous avons évalué les performances du modèle proposé et nous avons montré qu'il améliore de 10% le coût de calcul.

Le problème sur lequel on travaille actuellement et le déploiement de ce modèle dans un cas pratique. Va-t-on procéder d'une manière centralisée ou bien distribuée?

III. TRAVAUX REALISES

Dans le cadre de ce projet de thèse, nous avons soumis un article intitulé "Towards authenticated network coding for named data networking" dans la conférence softcom 2017. Le papier a été accepté et la conférence aura lieu le 21-23/09/2017.

Nous travaillons actuellement sur le déploiement du modèle proposé dans un cadre pratique. Nous étudions le cas d'un déploiement centralisé en utilisant les SDNs (Software-Defined Networking). Un papier est en cours de rédaction sur ce volet. Il s'agit de "Authenticated Network coding for Software-Defined named data networking".

Deux PFE ont été proposés l'année passée sur le sujet également, il s'agit de :

- Le streaming multimédia dans les réseaux centrés contenu mobiles. Les étudiants ont proposé une stratégie collaborative P2P utilisant le codage vidéo scalable. Ils ont également montré que le système proposé améliore considérablement les performances du streaming comparé à un système de streaming traditionnel. J'ai commencé également la rédaction d'un papier sur ce sujet. Il s'agit de " Peer-to-peer collaborative video-on-demand streaming over mobile content centric networking".

- Les réseaux centrés contenu pour les applications de soin de la santé intelligentes. L'étudiant a proposé dans ce travail

un schéma d'attribution des noms avec des stratégies de transmission des requêtes. Ce travail est en cours de réalisation et l'étudiant va soutenir bientôt. J'envisage de rédiger un papier sur ce thème également.

IV. PERSPECTIVE

J'envisage dans la suite de ce travail de réaliser les tâches suivantes:

- Déployer le modèle d'optimisation proposé en utilisant un algorithme distribuée. Ce travail fera l'objet d'un article que je vais soumettre dans une revue.

- Publier les deux articles sur les PFEs proposés l'année passée. A savoir le streaming vidéo et les services de soins de la santé.

- Proposer un nouveau sujet PFE sur l'application du codage réseau avec les techniques de streaming dans les réseaux centrés contenu et étudier le problème de l'évolution de la politique d'accès.

V. CONCLUSION

Présentés pour la première fois par Van Jacobson en 2006, les réseaux centrés contenu ont rapidement gagné l'intérêt d'une large communauté scientifique. Leur principe de fonctionnement consiste à permettre la recherche directe d'un contenu sans avoir à identifier son détenteur. Dans ce travail nous avons proposé une solution qui permet d'assurer un bon compromis entre le coût de calcul et le niveau de sécurité dans un CCN utilisant le codage réseau. Nous avons également travaillé sur un contenu particulier du web, il s'agit du streaming multimédia et nous avons proposé un système de streaming performant. Nous envisageons de déployer le modèle proposé dans un scénario centralisé et un scénario distribué au cours de cette année.

References

- [1] V. Jacobson, D. Smetters, J. Thornton, M. Plass, N. Briggs, and R. Braynard, "Networking named content", in CoNext 2009.
- [2] Seung-Hoon Lee, Mario Gerla, Hugo Krawczyk, Kang-Won Lee, and Elizabeth A Quaglia. Performance evaluation of secure network coding using homomorphic signature. In IEEE International Symposium on Network Coding (NetCod) , 2011.
- [3] R. Koetter and M. Médard, "An algebraic approach to network coding," IEEE/ACM Trans. Netw., vol. 11, no. 5, pp. 782–795, Oct. 2003.
- [4] <https://www.parc.com/work/focus-area/content-centric-networking/>
- [5] S. Agrawal, D. Boneh, X. Boyen, D. Freeman. Preventing Pollution Attacks in Multi-source Network Coding. In PKC'10, LNCS 6056, pp. 161–176, 2010.
- [6] Nuttapon Attrapadung and Benoit Libert. Homomorphic network coding signatures in the standard model. In Public Key Cryptography - PKC 2011, volume 6571, page 17, 2011.

Social Computing for Enhancing Collective Intelligence

Nadia BENAHMED EL ALLIA

Directeurs de thèse : Abdessamed Réda GHOMARI

Résumé — Les organisations doivent adopter et s'adapter à tous les changements technologiques issus du web tout en maintenant ou améliorant ses objectifs. Elles utilisent l'intelligence collective grâce aux connaissances et aux compétences qui sont disséminées chez des acteurs internes et externes pour innover de nouveaux produits ou services. Une nouvelle génération dite technologie sociale accroît la capacité des organisations à tirer parti de l'intelligence collective. L'objet de ce travail de recherche est d'analyser et questionner le social computing en particulier les technologies sociales et leur capacité à rehausser l'intelligence collective pour une organisation 2.0 plus particulièrement les écoles d'ingénieurs ouverts sur leur environnement socio-économique.

INTRODUCTION

Les interactions à tous les niveaux apparaissent comme une réelle force pour les organisations agiles, à partir du moment où des idées nouvelles, produits, services voire solutions émergent des multiples échanges et plus particulièrement les processus business. Les interactions au même titre que l'environnement externe, sont les moteurs de l'innovation et du changement dans des organisations agiles. Les organisations s'interrogent d'ailleurs de plus en plus sur les apports que peuvent procurer les modèles ainsi que les principes du Web collaboratif grand public (wiki, réseaux sociaux,...) voire du social computing de façon générale.

Au cours de ces dernières décennies, une nouvelle génération de technologies issue du web a apparu dans le monde de l'entreprise. Cette génération est passée de la version 1.0 vers la version 2.0 qui se concentre principalement sur les acteurs sociaux (acteurs, contributeurs et collaborateurs) [1]. Ces organisations doivent donc adopter et s'adapter à tous ces changements tout en maintenant ou améliorant ses objectifs. L'organisation doit donc encourager l'innovation et augmenter ces recettes avec ces outils permettant ainsi d'améliorer la productivité et de favoriser la collaboration.

Les usages collaboratifs, l'intégration des outils du Web 2.0, la valorisation de l'intelligence collective avec toutes les parties prenantes de l'organisation peuvent aider cette dernière à atteindre une meilleure performance économique et maintenir un avantage compétitif. Cette nouvelle génération dite technologie sociale accroît la capacité des organisations à tirer parti de l'intelligence collective [2].

La connaissance et l'expertise sont disséminées chez des personnes se trouvant aussi bien à l'intérieur qu'en dehors de limites formelles de l'organisation. La mise en commun des connaissances collectives et leur libre circulation au sein de l'organisation sont vues comme une prémisses à l'émergence de l'intelligence collective.

Les universités et les autres écoles d'ingénieurs sont des organisations sociales et travaillent à exploiter leur intelligence collective grâce à l'acquisition, au stockage, au partage et à l'évolution de la connaissance [3] liés à leur mission principale: (i) enseignement, (ii) recherche

scientifique et (iii) économique et influence sociale. Cette troisième mission axée sur l'innovation, l'esprit d'entrepreneuriat et l'engagement social est considérée comme complémentaire aux deux missions précédentes et tire parti d'un contexte favorable à l'ouverture à l'environnement socioéconomique, à la coopération internationale et aux incitations pour le développement de l'esprit d'entrepreneuriat [4], [5], [6].

Le présent document est structuré comme suit: tout d'abord, nous présentons la problématique de notre recherche (section II) suivie par les travaux réalisés (section III). La dernière section sera consacrée aux perspectives à savoir le travail qui reste à faire.

PROBLEMATIQUE

A mesure que les organisations manifestent plus d'intérêt à mettre à profit les connaissances détenues par leurs employés, les clients ainsi que les collaborateurs internes ou externes, le concept d'intelligence collective s'imposera de plus en plus en tant que mécanisme permettant d'inciter ces personnes à identifier de nouvelles opportunités à forte valeur ajoutée, résoudre des problèmes difficiles et de mettre en œuvre rapidement des innovations intéressantes.

Notre travail de recherche consiste à questionner et à analyser le potentiel du "social computing" et sa capacité à rehausser l'intelligence collective dans une organisation agile.

Nous nous focalisons particulièrement sur les grandes écoles d'ingénieurs algériennes que nous considérons comme des organisations agiles. Les deux questions de recherche subsidiaires qui se posent sont:

- Pourquoi et comment est-il possible d'appréhender les écoles d'ingénieur comme un système d'intelligence collective ?
- Quelles passerelles mettre en place entre les savoirs et connaissances intra-école et issues de l'écosystème ?

TRAVAUX REALISES

Plusieurs thématiques touchent notre sujet de recherche à savoir: l'organisation 2.0 [1] [7] (voir les institutions académiques de types écoles d'ingénieurs), l'intelligence collective [8] [9] [10] [11] [12] [13], le social computing et particulièrement le web social [14] [15].

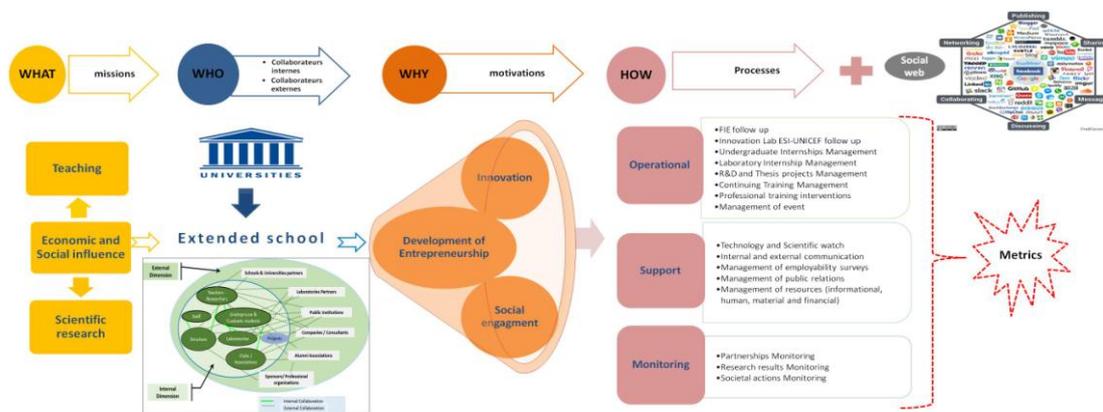
Nous avons publié un premier papier à "First International Conference on Business Intelligence and Applications" [16] afin de montrer comment améliorer l'intelligence collective dans une organisation de forme globale en exploitant le potentiel du web 2.0.

Une thématique complémentaire a été étudiée qu'est le capital intellectuel [17]. Ce dernier correspond à la connaissance intangible dont le capital relationnel génère des relations vers l'extérieur. La littérature présente un intérêt croissant sur le capital intellectuel au niveau des universités.

Pour mener à bien la recherche, les chercheurs utilisent la recherche empirique [18] [19] [20] [21].

Nous avons réalisé un méta modèle représentant le système d'intelligence collective tout en exploitant le potentiel des technologies du social web appliqué aux écoles d'ingénieurs algériennes. Nous avons donc soumis ce méta modèle à la conférence "International Conference on Mathematics and Information Technology" (en attente de la réponse). Ces écoles de savoirs considérées comme des écoles étendues sont devenues ouvertes à leurs écosystèmes. Une nouvelle mission

leur a été donc attribuée complémentaires aux missions élémentaires (enseignement et recherche scientifique). Cette mission représentée par l'innovation, l'esprit d'entrepreneuriat et le développement technologique a été instaurée par un décret publié dans le journal officiel (voir décret n° 16-176 du JO 36 en date du 19/06/2016). Notre étude de cas est l'école supérieure d'informatique où nous avons identifié un ensemble de processus dont nous avons aussi identifié une liste d'indicateurs.



Méta modèle d'intelligence collective sur le cas d'une école d'ingénieurs: Cas ESI

PERSPECTIVES

Nous envisageons dans la suite de notre travail d'adopter une approche de recherche empirique à travers des études de cas [22] :

- Analyser le potentiel technologique du web social à base de questionnaires ciblés;
- Étendre l'étude empirique au réseau des écoles d'ingénieurs;
- Modéliser et implémenter les processus génériques pour toute école d'ingénieurs;
- Développer des plates-formes collaboratives support pour soutenir les processus génériques de réseaux d'écoles étendues.

REFERENCES

[1] Boughzala I. and DE Vreede G. Vers l'organisation 2.0: un nouveau modèle basé sur l'intelligence collective. AIM'10, 15^{ème} Colloque Association Information et Management, La Rochelle, France, pp. 1-10, 2010.

[2] Lesser E., Ransom D., Shah R. and Pulver B. Intelligence Collective: Capitaliser sur le talent de la foule. IBM Institute for Business Value. 2012.

[3] Nonaka I. and Takeuchi H. The Knowledge-Creating Company. Oxford: University Press, 1995.

[4] Abreu M., Demirel P., Grinevich V. and Karatas-Ozkan M. Entrepreneurial practices in research-intensive and teaching-led universities. Small Business Economics, n°47, pp.695-717, 2016.

[5] Foss L. and Gibson D.V. The Entrepreneurial University: Context and Institutional Change. Routledge, 2015.

[6] Bharat R. and Bala M. The role of Universities in Encouraging Growth of Technology-Based New Ventures. International Journal of Innovation and Technology Management, Vol. 14, n° 4, 2016.

[7] Super C., Naula V. and Nyffeler N. Entreprise 2.0 : START-UP & PME. Livre Blanc, 30/08/2013.

[8] Zaïbet Gréselle O. Définir et repérer l'intelligence collective dans les équipes de travail opérationnelles : Le cas d'une PME du secteur de l'électronique, Management & Avenir 4/2007 n°14, pp. 41-59, 2007.

[9] Malone T.W. and Klein M. Harnessing Collective Intelligence to Address Global Climate Change. Innovation: Technology, Governance, Globalization, Vol2, n°3, pp. 15-26, 2007.

[10] Malone T.W., Laubacher R. and Dellarocas C. The collective intelligence genome. MIT Sloan Management Review, Vol. 51, n° 3, pp. 21-31, 2010.

[11] Marcin Maleszka and Ngoc Thanh Nguyen. Integration computing and collective intelligence. Expert Systems with Applications, Expert Systems with Applications 42 (2015) 332–340, Elsevier.

[12] Noubel J.F. Intelligence Collective: la révolution invisible. The Transitioner, publié en novembre 2004, révisé en août 2007

[13] Lévy P. L'intelligence collective: Pour une anthropologie du cyberspace. La découverte, 1997.

[14] Chui M. A. Miller and Roberts R.P. Six ways to make web 2.0 work, The McKinsey Quarterly, in Business Technology, 2009.

[15] O'reilly T. and Battelle J. Web Squared: Web 2.0 five years on. Web 2.0 Summit, San Francisco, CA, (2009).

[16] Andriessen D. Making Sense of intellectual capital, Designing a Method for the Valuation of Intangibles. Edition Elsevier Butterworth Heinemann, USA, 2004.

[17] Benahmed El Allia N. et Ghomari A.R. Enterprise 2.0: How to Enhance Collective Intelligence? Web social potential. The 2016 International Conference on Business Intelligence and Applications (ICBIA'16), pp. 26-31, March 1-3, 2016

[18] Cabrita M.R, Secundo G., Passiente G. and Matos F. Intellectual Capital in Academic Entrepreneurship: Moving Beyond Measurement, 16th European Conference on Knowledge Management University of Udine, Italy, 3-4 September 2015.

[19] Bezhan I. Intellectual capital reporting at UK universities. Journal of Intellectual Capital, vol11, n°2, pp. 172-207, 2010.

[20] Trencher G., Yarime M., McCormick Kes B., Dolle C. N. H. and Kraines S. B. Beyond the third mission: Exploring the emerging university function of co-creation for sustainability. Science and Public Policy 41, pp. 151-179, 2014.

[21] Secundo G., Dumay J., Schiuma G. and Passiante G. Managing intellectual capital through a collective intelligence approach: An integrated framework for universities. Journal of Intellectual Capital, Vol. 17, n° 2, pp. 298-319, 2016.

[22] Robert K. Yin. Case Study Research Design and Methods (5th ed.) 2014. Thousand Oaks, CA: Sage. 282 pages.

Optimal Deployment and Setting of Wireless Sensor Nodes for Energy Management in Smart Buildings

Laidi Roufaida

Supervisor : Prof. Djenouri Djamel

Team : WSN & Applications group, CERIST

Abstract—This report shows the achievement accomplished during the year of 2016/2017 in the frame of a PhD thesis treating the optimal sensors deployment for the case of smart buildings.

I. INTRODUCTION

The humanity has witnessed during the last few years constant economic growth, large proliferation of electronic and consumer appliances, which inevitably cause sharp increase of energy consumption in residential and commercial buildings worldwide. This increase has dramatic footprints on the population, environment, and the economy. Green and smart building strategies will play a pivotal role to reduce this footprint and to maximize economic and environmental performance.

The use of modern Information and Communication Technologies (ICT), such as wireless sensor networks, for building control, is one of the promising strategies for the future. Combining potential real time information from the sensors with reliable predictions, efficient automation policies are expected to be achieved, such as controlling lighting, heating, air conditioning, etc. according to contextual and climate conditions, e.g., room occupancy, sunlight, ambient temperature, etc., which will significantly reduce the energy consumption.

II. PROBLEM STATEMENT

The aim of this thesis is to (after a deep study and taxonomy on the state-of-the-art) focus on the node deployment problem. This problem is NP-hard even with simplified assumptions. The deployment in this project is not limited to determine node positions, but also appropriate setting of operational parameters such as cycle durations, timeouts, etc. It is also requested to define ways to effectively set these parameters using stochastic models.

This thesis makes part of an ongoing project at CERIST that aims at proposing solutions based on sensors and Internet of Things technologies for energy management in smart building¹.

III. ACHIEVED WORK

Information on space occupancy is vital for smart building applications and services, such as energy management and surveillance systems. It steers service provision with respect to temporal and spacial constraint, e.g. when/where actuating lighting system, HVAC, for how long/much, etc. The accuracy of detection depends directly on the position and number of the sensors to be deployed. For that, our first contribution

concerns optimal motion sensor deployment that increases sensing coverage while reducing the economic cost. Scalability to large surfaces is also targeted.

A. Related work

This section discuss solutions treating motion sensors deployment in buildings and aim at increasing coverage while reducing costs. [1] targets detection and location in green buildings. The authors divide the area into uniform grids and assume that not all grid points are suitable for deployment because of building's properties, without mentioning their method of selecting suitable deployment locations. A binary vector encoding is proposed, an element is set at 1 if a sensor is deployed in the corresponding grid point. The length of the vector is the number of grid points, which may make the solution not suitable for large scales. The problem is resolved using a Particle Swarm Optimization (PSO) approach.

[2], [3] are two propositions that focus on sensors sensing area. [2] investigates the effect of obstacles (internal walls, doors, closets ect.) on sensing area, while [3] considers the gaps of PIRs' field of view. Sensing gapes due to PIRs lenses makes them prone to false negatives. Considering them while selecting nodes locations significantly reduces the amount of false negatives, but makes the solution impractical since detections zones change depending on the PIR type and ceiling hight. Both works use Integer Linear Programming (ILP) to resolve their problems. Most deployment problems formulations are NP-hard or NP-complete [4], using exact methods to resolve them is not suitable.

[5], [6] highlight the fact that not all zones in a building have the same importance in term of occupancy. They extract zones importances using furnitures locations which are introduced manually. Path Finding Algorithms(PFA) are then used between furnitures to simulate occupants trajectories, these paths are used to build a floorplan heatmap. The area is presented as regular grids and the importance of a grid point depends on the heatmap. [5] resolve the problem using a greedy algorithm, while[6] uses a Genetic Algorithm.

B. Contribution

In a building area, different zones are not occupied equally. Similarly, the targeted coverage in the monitored area should be differentiated from an area to another, and the sensor placement should be tuned accordingly. This allows to increase accuracy in coverage by concentrating on the appropriate regions, which reduces the deployment cost. The importance of a zone is proportional to the time the users spend in it, which depends upon the furniture, users' habits and preferences. To

¹http://wsn.cerist.dz/?page_id=566

our knowledge, the only works that consider furniture when defining the zone's importance are [3], [6], [5]. While furniture information is important and effects zones occupancy, it is not sufficient as it does not describe the time spent in a zone. For example, in a single-occupant office, it is possible to have two chairs, the main one and the one for guests. Solutions based on furniture do not allow to make any distinction between the areas where the chairs are located, while the former is largely more occupied than the latter. Further, existing solutions rely on manual introduction of furniture locations and types, which is impractical and prone to mistakes.

Existing works on sensor deployment in buildings, such as [1], [2] discretize the deployment area into uniform grids. The granularity of the grid has balancing impact between accuracy and computation time (scalability), a small distance between grid points increases accuracy but also the computation time, which reduces scalability, and vice-versa. Further, most works in literature requires optimal setting for complex parameters e.g. a_i minimum number of sensors that needs to cover point i ([7]), c_{max} maximum preferable coverage score per grid point([5]), Genetic Algorithm parameters ([6]) .ect, which needs deep technical knowledge on the deployment algorithm and some pre-deployment tests. To overcome these drawbacks, our goals in this work are,

- Increasing coverage while reducing cost: For this, we define areas priorities proportionally to the time spent by the occupants. A learning phase is needed to collect occupants behavior.
- Increasing scalability without compromising accuracy: For that, the proposed solution concentrates the possible deployment points around areas of high importance, instead of using uniform distribution.
- Reducing the number of required parameters: we use only two parameters, i) the budget, and, ii) the type of sensors to be used. These parameters are naturally understood and can be easily tuned by a user without the need of deep technical knowledge of the solution.

In order to find optimal nodes coordinates, the solution proceeds in steps. 1) Detect zones occupation time during a learning period where ground truth data about users locations is collected. 2) Generate an undirected weighted graph structure based on the collected data. Graph vertexes coordinates are possible deployment points and vertexes spreading across a zone relays on the zone's importance. Graph edges describe sensors coverage: two vertexes i, j are connected, if deploying a sensor on i gives a non-zero probability to cover j . 3) maximizing the coverage with a limited number of sensors is formulated as "the maximization of a submodular function with cardinality constrains" problem, which is solved using a greedy algorithm.

Simulation results show an improvement in term of coverage and computation time comparing to manual deployment and Vlasenko et al. [5] solution .

IV. PERSPECTIVES

For the next steps we aim:

- Finalize and submit the current work.

- Proof of concept by real deployment.
- Improve battery usage.

V. CONCLUSION

The advancement accomplished during last year correspond to the initial time line.

REFERENCES

- [1] L. Wang, X. Zou, Q. Meng, and X. Song, "An optimal strategy for the deployment of sensor nodes in green buildings," in *Intelligent Control and Information Processing (ICICIP), 2015 Sixth International Conference on*, pp. 209–213, IEEE, 2015.
- [2] M. P. Fanti, M. Roccotelli, J. J. Lesage, and G. Faraut, "Motion detector placement optimization in smart homes for inhabitant location tracking," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1–6, Sept 2016.
- [3] A. Ouadjaout, N. Lasla, D. Djenouri, and C. Zizoua, "On the effect of sensing-holes in pir-based occupancy detection systems," in *SEN-SORNETS 2016 - Proceedings of the 5th International Conference on Sensor Networks, Rome, Italy, February 19-21, 2016.*, pp. 175–180, 2016.
- [4] D. S. Deif and Y. Gadallah, "Classification of wireless sensor networks deployment techniques," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 834–855, 2014.
- [5] I. Vlasenko, I. Nikolaidis, and E. Stroulia, "The smart-condo: Optimizing sensor placement for indoor localization," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 3, pp. 436–453, 2015.
- [6] D. Yang, K. Rao, B. Xu, and W. Sheng, "Pir sensors deployment with the accessible priority in smart home using genetic algorithm," *Int. J. Distrib. Sen. Netw.*, vol. 2015, pp. 2:2–2:2, Jan. 2015.
- [7] B. Wang, J. Zhu, L. T. Yang, and Y. Mo, "Sensor density for confident information coverage in randomly deployed sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 3238–3250, May 2016.

Intégration spatio-temporelle de données et connaissances pour la prévention des risques professionnels en santé

HAROUN Hayat

Directeurs de thèse : GHOMARI Abdessamed Réda

Équipe : MSI

Résumé—La prévention des risques professionnels ambitionne d'écarter les risques et/ou leurs sources, et d'améliorer les conditions d'hygiène et santé au travail. Notre travail de recherche consiste à questionner les méthodes et techniques et leur capacité de fournir une traçabilité des données sur l'exposition aux risques des travailleurs aussi bien en mobilité spatiale que temporelle.

I. INTRODUCTION

Assurer la santé et la sécurité des salariés, anticiper leur usure professionnelle et diminuer leurs pénibilités physiques et morales constituent des éléments de performance pour l'entreprise. Un salarié en bonne santé augmente ainsi son efficacité au service de l'entreprise.

La prévention des risques professionnels est l'ensemble des mesures prises pour empêcher un risque de survenir. La politique de la prévention des accidents et des maladies professionnelles reste insuffisante en Algérie qui enregistre annuellement et ce, depuis cinq ans, 600 décès sur les 50 000 accidents de travail recensés [1].

Ce nombre important de victimes nécessite l'étude de ce phénomène pour en extraire tous les facteurs d'influence et permettre ainsi de proposer des mesures préventives. La plupart des études actuelles sous-emploient l'information temporelle et géographique ce qui produit une analyse partielle voire incomplète de ce phénomène.

Ce travail de recherche ambitionne de réfléchir à la problématique d'intégration aussi bien l'axe spatial que l'axe temporel dans l'analyse afin d'améliorer l'exploitation et la visualisation des données de prévention par les médecins du travail. Pour cela, nous allons faire recours aux systèmes dits « Géo-Décisionnels ».

II. PROBLÉMATIQUE

Pour élaborer des politiques de prévention, il est important voire impératif de disposer des informations de synthèse sur l'état actuel, c'est-à-dire un constat de la situation présente sur lequel des mesures correctives et préventives vont être émises. Donc nous allons aborder les problèmes liés à l'entrepôt des données en un premier temps. Par la suite, il faut traiter l'intégration des deux dimensions : spatiale et temporelle au sein de l'entrepôt de données. Enfin, et afin d'explorer et d'analyser ces masses de données selon les deux axes spatial et temporel, nous allons nous référer aux applications SOLAP (Spatial On-Line Analytical Processing).

III. TRAVAUX RÉALISÉS

Cette première année de thèse a été consacrée à faire un état des lieux de la prévention des risques professionnels. Nous nous sommes focalisés sur le contexte algérien, et ce pour connaître davantage ce qui se fait pour l'instant, les manques et faiblesses et ainsi agir de manière efficace à cette situation tout en essayant, toujours, de voir en quoi l'intégration des données spatiotemporelles pourrait être bénéfique à l'amélioration de la prévention des risques. Dans cette optique, nous avons entamé des discussions sur les pistes de collaboration avec le service de médecine du travail (SUMT) de l'hôpital de Rouiba. En effet, nous avons assisté à maintes reprises à des colloques établis par le service par le biais desquels nous avons pu analyser leur processus du travail et pu recenser le besoin à un tel projet de recherche et les retombées d'une éventuelle collaboration à ce sujet.

Par ailleurs, nous avons élaboré une synthèse des travaux relatifs au domaine Géo-décisionnel. Nous avons tenté de faire le tour d'horizon de ce domaine : ses fondements, ses avantages, et ses domaines d'application,...etc.

A. La prévention des risques professionnels en Algérie

La prévention des risques professionnels est, avant tout, une affaire interne de l'organisme employeur. Il est de sa responsabilité d'offrir à ses travailleurs un milieu de travail sûr et salubre en conformité avec les normes légales et réglementaires émises par les pouvoirs publics. Il en relève des responsabilités diverses aussi bien sur le plan interne (Commission paritaire d'hygiène et de sécurité, Service d'hygiène et de sécurité, Structure de médecine du travail) que sur le plan externe (notamment l'inspection du travail et les organismes de prévention). Tous ces acteurs interagissent dans un processus commun et itératif (figure 1) et peuvent être classés comme illustré dans la figure 2 :

B. Intégration des données spatiotemporelles

L'intégration des données spatiotemporelles s'avère nécessaire dans le processus de prévention. En effet, l'exposition aux risques professionnels est sujet de ces deux dimensions. Tracer la mobilité du travailleur dans le temps exige de suivre l'évolution de son exposition à des facteurs ou agents à risque. C'est ce qui permettra d'élaborer un suivi médical concis. A titre d'exemple l'exposition à l'amiante peut engendrer des maladies professionnelles qui ne se manifestent pas forcément durant l'exposition mais plutôt en post-exposition. Donc, omettre la dimension temps, mène à



Fig. 1: processus de prévention.

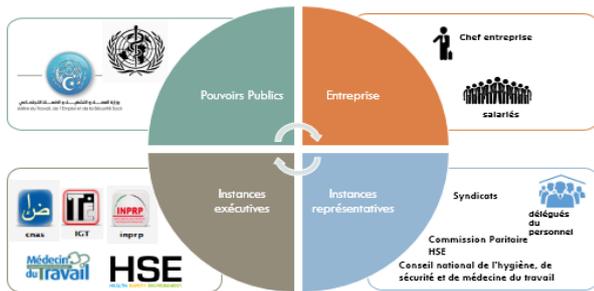


Fig. 2: Acteurs de la prévention.

confusion sur l'origine de la manifestation d'une maladie dite « professionnelle » dans ce cas.

Par ailleurs, les salariés sont souvent amenés à se déplacer dans un espace déterminé qui correspond à leur poste de travail. Le déplacement peut avoir lieu dans le même bureau, la même entreprise ou plus largement entre entreprises.

L'exposition moyenne d'un salarié à son poste de travail est, par conséquent, le résultat d'une succession d'expositions variables dans le temps et dans l'espace.

C. Les systèmes géo-décisionnels

Le géo-décisionnel désigne à la fois concept et technologie permettant de combiner l'informatique décisionnelle (BI) et les systèmes d'information géographique. Le terme Spatial OLAP désigne aussi une nouvelle famille d'outils s'appuyant sur un enrichissement de la modélisation multidimensionnelle de manière à prendre en compte non seulement le caractère multidimensionnel des données mais aussi le caractère spatial ou géographique. Il s'agit d'un concept technologique mis au point par l'équipe du Pr. Yvan Bédard au Centre de Recherche en Géomatique (CRG) de l'Université Laval [2].

Depuis son apparition en 1990, cette technologie ne cesse d'évoluer à la fois dans l'enrichissement de ses fondements mais aussi dans ses domaines d'application qui en témoignent l'intérêt (notamment la foresterie[3], le trafic routier [4], les accidents [5], les risques environnementaux[6], [7], santé publiques[8], [9], [10],...etc.)

IV. PERSPECTIVES

Notre objectif à présent, est de faire une étude de faisabilité d'une telle application des SOLAP à notre contexte de prévention des risques professionnels. Le but étant de développer une approche méthodologique spatio-temporelle utile à la production d'une cartographie des risques qui permettra par la suite d'analyser l'occurrence des accidents de travail et d'identifier des indicateurs qui, quant à eux, décriront la répartition des maladies professionnelles selon les entreprises, les domaines d'activité, ...etc. C'est-à-dire une analyse multi-dimensionnelle et spécialement temporelle et spatiale.

V. CONCLUSION

Cette première phase de recherche nous a permis de bien cerner le problème posé. Nous avons pu nous positionner dans notre état de l'art. Nous avons opter pour piste de recherche les travaux d'application du domaines du GéoDécisionnel. Il est sujet maintenant de définir une solution potentielle au problème, qui à ce stade, a pour prémisses l'application de la méthodologie et technologie SOLAP qu'il faudra valider, par une mise en oeuvre d'un outil intégrant les dimensions spatiale et temporelle et évaluer par la suite l'intérêt d'une telle démarche sur l'amélioration de la prévention des risques professionnels.

RÉFÉRENCES

- [1] Hygiène, santé et sécurité du travail : la situation est alarmante en Algérie, <http://www.algerie-focus.com/2013/08/hygiene-sante-et-securite-du-travail-la-situation-est-alarmante-en-algerie/>
- [2] Y. Bédard, "Le géodécisionnel origine évolution état de l'art enjeux R&D," Sophia Antipolis, France, octobre, vol. 29, 2010.
- [3] M. Miquel, Y. Bédard, et A. Brisebois, "Conception d'entrepôts de données géospatiales à partir de sources hétérogènes. Exemple d'application en foresterie," Ingénierie des Systèmes d'information, vol. 7, no 3, p. 89 111, 2002.
- [4] Y. Bédard, " Application de la géomatique aux permis spéciaux pour le transport de chargement indivisible et le transport des marchandises dangereuses," 2006.
- [5] D. A. Khalissa, I. Frihi, K. Boukhalfa, et Z. Alimazighi, "De la Conception d'un Entrepôt de Données Spatiales à un Outil Géo-Décisionnel pour une Meilleure Analyse du Risque Routier," in INFORSID, 2013, p. 181 196.
- [6] B.H. Tran, C. Plumejeaud-Perreau, A. Bouju, V. Bretagnolle, "Modélisation et analyse de données environnementales à travers une ontologie spatiotemporelle," Atelier INOVIVE 4ième édition, Juin 2016, Montpellier, France.
- [7] J. Iris, "Contribution de la méthodologie et de la technologie géodécisionnelle pour l'aide à l'évaluation des risques naturels dans le secteur de l'assurance en France," Sciences de l'ingénieur [physics], Ecole Nationale Supérieure des Mines de Paris, 2009.
- [8] Y. Bédard, P.Gosselin, S. Rivest, M.J. Proulx, M. Nadeau, G. Lebel, and M.F Gagnon "Integrating GIS components with knowledge discovery technology for environmental health decision support," International journal of medical informatics, vol. 70, no 1, p. 79 94, 2003.
- [9] B. Parmanto, M. V. Paramita, W. Sugiantara, G. Pramana, M. Scotch, and D. S. Burke, "Spatial and multidimensional visualization of Indonesia's village health statistics," International journal of health geographics, vol. 7, no 1, p. 30, 2008.
- [10] Réseau Convergence Network, "Analyse spatio-temporelle des maladies d'origine alimentaire," Réseau Convergence. [En ligne]. Disponible sur : <http://www.convergence.ulaval.ca/chercheurs/analyse-spatio-temporelles-maladies-dorigine-alimentaire/>. [Consulté le : 11-avr-2017].

Conception et génération d'un Meta-Modèle de connaissances à base de trace pour l'ingénierie des EIAH

Samia AIT ADDA

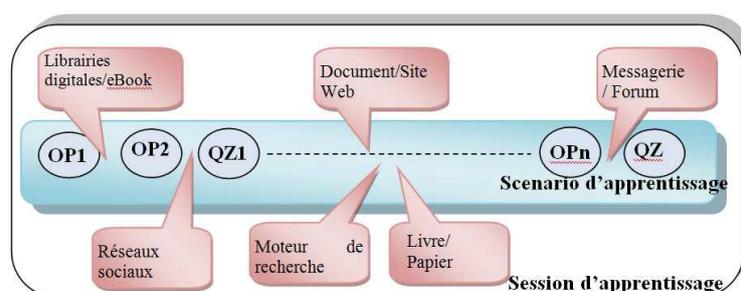
Directeurs de thèse : Amar BALLA

Équipe : (CoDesign, Optimisation, MSI, EIAH, Image, SURES)

Résumé— Ce papier a pour objectif de présenter l'état d'avancement à travers les travaux de recherche, que nous avons menés dans le cadre d'une thèse, et ce concernant la modélisation de l'apprenant et l'analyse de son comportement à base de traces collectées durant son apprentissage. La raison principale de cette analyse est d'assister le tuteur pour reconnaître les apprenants en difficulté, et d'aider l'auteur du cours d'identifier les parties du cours nécessitant une restructuration et/ou d'être plus détaillé. Comme précision, ce travail a abouti à des résultats satisfaisants qui ont fait l'objet d'un article revue et la rédaction d'un manuscrit de thèse à faire valider

Mots-clés. EIAH, Modélisation de l'apprenant, Analyse des traces, Ontologie du domaine, Indexation conceptuelle, Adaptabilité

tout type d'artefact qui puisse aider l'apprenant à évoluer et à comprendre comme l'explique si bien la figure ci-dessous.



OP : Objet Pédagogique

QZ : Quiz et test

Fig. 1. Milieu d'apprentissage

I. INTRODUCTION

Le terme utilisé actuellement pour désigner les systèmes construits est EIAH pour Environnements Informatiques pour l'Apprentissage Humain. Un EIAH est un environnement informatique conçu dans le but de favoriser l'apprentissage humain, c'est-à-dire la construction de connaissances chez un apprenant [1]. Depuis quelques années, l'environnement informatique pour l'apprentissage humain s'est rapidement évolué. Ce mode d'enseignement doit son développement à l'évolution des technologies de l'information et de communication et de ces usages.

Toutefois, avec le nombre croissant d'utilisateurs de ces systèmes et la diversification de l'espace et du temps entre les apprenants et les enseignants, nous assistons à plus d'exigence d'adaptabilité, de flexibilité et d'efficacité de l'apprentissage, ce qui soulève des problèmes d'une grande diversité, essentiellement la personnalisation et l'adaptabilité d'apprentissage [2].

En effet, détecter quelles sont les connaissances du domaine qui posent problème peut permettre à un EIAH de proposer les activités pédagogiques les plus appropriées à ces utilisateurs. Pour cela, les exigences précédemment citées doivent prendre en compte les caractéristiques des usagers issus de ces interactions avec les différents outils des systèmes qui contribuent à l'apprentissage.

Avec l'avènement et explosion du Web 2.0, le milieu d'apprentissage actuel n'est plus restreint uniquement aux plateformes d'apprentissage et ses différents services, mais à

II. PROBLÉMATIQUE

Mais avant d'aller plus loin, il convient d'examiner un certain nombre d'aspects en rapport avec cet accès à des ressources externes. Cependant notre problématique de recherche s'agit ainsi de se demander pourquoi l'apprenant visite (recherche, partage et communique) sur des contenus externes au système d'apprentissage durant sa session d'apprentissage? Est-ce par manque de concentration, de motivation ou de compréhension? Les contenus visités (recherchés, partagés et communiqués) sont-ils en rapport avec le domaine étudié? Si oui, à quel degré? Et sur quels concepts du domaine portent ces contenus?

III. CONTRIBUTION

Pour détecter les connaissances du domaine supposé mal assimilé par les apprenants et/ou mal conçu par les concepteurs, ainsi que les apprenants en difficulté, nous proposons alors à un Méta-modèle de connaissance d'un apprenant modélisé grâce à une ontologie du domaine. Ce modèle contient : (1) le modèle du domaine, (2) le modèle d'interaction, (3) le modèle Pédagogique, (4) et le modèle de connaissance d'un apprenant. Une première finalité de ce modèle est de permettre l'adaptabilité et la personnalisation de l'apprentissage. Par conséquent des approches ont été alors découlées et ce, selon le type d'interaction de l'apprenant durant son apprentissage (voir figure 1). Nous avons divisé ces approches sur deux types, trois approches concernent les interactions de l'apprenant au sein de la plateforme (cours, discussion et requêtes sur les librairies digitales), deux autres approches sur les interactions des apprenants en dehors de la

plateforme (consultation des pages Web, et requêtes sur les moteurs de recherche)

Nous allons décrire dans ce qui suit, les différents travaux ainsi aboutis pendant toute la durée d'élaboration de thèse et selon la chronologie du temps de leur réalisation :

1. La première approche que nous avons développée consiste en la gestion de l'expérience d'apprentissage à base de traces pour l'adaptabilité dans un EIAH [3].
2. La deuxième approche consiste à analyser les requêtes de l'apprenant sur les moteurs de recherche et d'identifier les mots-clés émis, que nous considérons par ailleurs comme les concepts du domaine qui posent problème à ces apprenants [4], [5].
3. La troisième approche, comprend l'analyse sémantique des messages édités par les apprenants, qu'il soit sur la messagerie, chat ou sur les forums de discussions, et ce dans le but de découvrir les concepts du domaine qui sont les plus discutés et communiqués par les apprenants, par conséquent, deviner les concepts problèmes ainsi que les apprenants en obstacle [6] [7],[8].
4. La quatrième approche quant à elle repose sur l'analyse conceptuelle des requêtes d'apprenants édités sur les catalogues des librairies digitales qu'elles soient internes à l'université ou à celle de la plateforme, ajouté à elle l'analyse du contenu consulté (résultat de la recherche sur catalogue) [9].
5. La cinquième approche consiste à analyser le parcours d'apprentissage de l'apprenant afin de détecter le comportement de désorientation et ceci grâce aux similarités sémantiques entre les déferentes ressources qui se suivent dans le parcours de l'apprenant [10], [11].
6. En finale, la sixième approche permet d'analyser les commentaires des apprenants postés sur les réseaux sociaux, plus précisément Facebook dans notre cas puisque c'est le réseau le plus utilisé en Algérie, et de détecter ainsi l'opinion de l'apprenant vis-à-vis du cours enseigné ainsi que les concepts du domaine mal assimilés et les apprenants en difficultés.

IV. CONCLUSION ET PERSPECTIVES

Les approches ainsi proposées permettent, comme il a été précisé dans partie contribution, de détecter les concepts du cours qui sont peu ou mal acquis par les apprenants et pointer ainsi les apprenants qui sollicitent ces concepts, qu'on considérera comme : des apprenants en difficultés et qui nécessitent du suivi et de l'accompagnement, tout en s'appuyant sur des paramètres désignés par les différents acteurs participants dans le processus d'apprentissage.

Des expérimentations ont été effectuées sur des groupes d'apprenants (182 étudiants de deuxième année du département informatique de l'université de Mouloud Mammeri de Tizi-Ouzou), sur un cours dont le domaine est modélisé à travers une ontologie du domaine en format SKOS (Simple Knowledge Organization System), afin de valider les approches et le modèle proposés. Les résultats ont été satisfaisants et nous ont permis d'enrichir encore le modèle de connaissance de l'apprenant pour contribuer davantage à

l'adaptabilité de l'apprentissage, la réussite des apprenants et essentiellement à l'efficacité de la formation.

Nous pouvons alors dire que les objectifs qui ont été assignés par la présente thèse ont été largement atteints et les travaux ainsi validés par les différents papiers conférences et revues nous ont permis ainsi de contribuer à développer de nouvelles méthodes pour modéliser et promouvoir le domaine de l'apprentissage en ligne. Le manuscrit final de thèse est rédigé, pour être évalué par des experts.

Nos perspectives visent à approfondir l'analyse des documents consultés par l'apprenant en dehors du système d'apprentissage tels que les blogs, forum et librairies digitales externes, afin de détecter leurs contenus sémantiques et corrélations, pour pouvoir encore préciser les manques du domaine qui ont poussé l'apprenant à consulter ces concepts du domaine depuis des ressources pédagogiques externes à la plateforme.

Références

- [1] Tchounikine, P. : Précis de recherche en ingénierie des EIAH. juin 2009 Disponible à <http://membresliglab.imag.fr/tchounikine/Articles/PrecisV1.pdf>.
- [2] Brusilovsky, P., Henze, N.: Open corpus adaptive educational hypermedia. In: Brusilovsky, P., Kobsa, A., Neidl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York (2007) 4-194.
- [3] S. Ait Adda, Balla, A. Faron Zucker, C. "Gestion de l'expérience d'apprentissage à base de traces pour l'adaptabilité dans un EIAH", 1^{er} Séminaire National sur les Technologies Educatives, SNTE 2012, 6 – 7 Mars 2012, Guelma, Algérie. http://labstic.net/snte/liste_articles.html
- [4] Ait Adda, S. ; Balla, A. Enrichment of learner profile through the semantic analysis of the learner's query on the Web, Innovative Computing Technology (INTECH), 2014 Fourth International Conference on DOI: 10.1109/INTECH.2014.6927773, IEEE (2014), Page(s): 7- 12
- [5] Ait Adda, S., Balla, A.: The Use of Ontology in Semantic Analysis of the Learner's Queries on the Web through Search Engines. IJWA 6(3): 121- 130 (2014)
- [6] Ait Adda, S. Balla, A. "Détection des difficultés d'apprentissage à travers l'analyse sémantique des messages " 3^{èmes} Journées Doctorales en Informatique de Guelma, Guelma, Algérie, 4 - 5 Décembre 2013.
- [7] Ait Adda, S., Balla, A.: The use of ontology in semantic analysis of the published learner's messages for adaptability. Proceedings of the 4th International Conference of Swarm Intelligence Based Optimization - ICSIBO 2014, LNCS, (2014) 106-114.
- [8] Ait Adda, S., Balla, A.: Enrichment of learner profile through a semantic analysis of the posted discussion, NNGT Int. J. on E-learning and Education, Vol. 1, July 2014, DOI : 01.IJEELE.2014.1.5, page(s) 1-7
- [9] . Ait Adda, S. ; Balla, A. The impact of digital library in detecting of learning difficulties, ISKO-Maghreb, 2014 4th International Symposium, DOI: 10.1109/ISKO-Maghreb.2014.7033449, IEEE, Page(s): 1- 6
- [10] Ait-Adda, S., Bousbia, N. Evaluation de la désorientation de l'apprenant dans un système d'apprentissage, Évaluation des Apprentissage et environnement Informatique, Agadir, 2015
- [11] Ait-Adda, S., Bousbia, N., Balla, A. A Semantic Analysis of the Learner's Disorientation, International Journal of Emerging Technologies in Learning, iJET – Volume 11, Issue 6, 2016, Page(s): 10- 18

Un système de recommandation des ressources Scientifiques

Khadidja BENMESSAOUD DIAB
Directeur de thèse : Dr. Faïçal AZOUAOU
Équipe : EIAH
Année d'inscription : 2015/2016

Résumé _ L'essor d'Internet a favorisé la mise en ligne de productions scientifiques sur le web, il s'agit en particulier d'articles scientifiques publiés par les chercheurs eux-mêmes ou par leurs institutions. Trouver alors une ressource scientifique pertinente devient une tâche difficile. Notre objectif est de proposer un système intelligent qui assiste les chercheurs à trouver les ressources scientifiques les plus pertinentes par rapport à leurs tâches courantes. Ce papier représente un aperçu des travaux de recherche réalisés durant ces deux années de thèse, et décrit brièvement notre proposition afin d'élaborer un système de recommandation basé sur l'apprentissage automatique.

I. INTRODUCTION

La première phase de la recherche est celle de la sélection d'informations pertinentes. En effet, tout chercheur a besoin de connaître précisément les travaux réalisés, à la fois pour s'y appuyer, mais aussi pour vérifier l'originalité de sa propre réflexion.

De nos jours, la publication et la diffusion des travaux de recherches, comme toute autre information, sont rendues facile et rapide grâce aux technologies web et mobiles. Malgré les avantages dus à cette évolution, d'autres problèmes se sont engendrés, notamment le fait que les chercheurs se trouvent submergés dans une grande masse d'information (infobésité) qui ne cesse de s'amplifier. C'est dans ce contexte que se situe notre travail.

Notre objectif est de proposer un système de recommandation ([1] et [2]) dédié aux chercheurs pour augmenter leur efficacité et leur rendement. Ce système recommande à chaque chercheur des ressources pertinentes, en particulier: les articles scientifiques ainsi que les chercheurs travaillant sur des thématiques similaires.

II. PROBLEMATIQUE

Étant donné que le volume de recherches publiées augmente de manière exponentielle. Il devient ainsi de plus en plus difficile pour un chercheur donné de repérer des articles pertinents conformément à ses besoins et activités courants. [3]

Il est également difficile pour un chercheur donné de trouver de chercheurs travaillant sur des thématiques similaires aux siennes, car une simple recherche sur le web lui propose un nombre très important de profils des autres chercheurs, ce qui rend le repérage des chercheurs de références difficile [4].

Tous ces éléments (productions scientifiques, profils de chercheurs) sont visibles dans le web, mais ils sont également de plus en plus présents dans les réseaux sociaux génériques (généralistes) [5] ou scientifiques [6], [7]. Ces derniers ont permis encore la création de nouvelles formes de collaboration, d'échange, de création et de partage des connaissances.

Il existe déjà plusieurs moteurs de recherche génériques [8], [9] ou scientifiques [10], [11]. Mais ces derniers ne sont

toujours pas adaptés aux besoins et profil de chaque chercheur. C'est pourquoi un système de recommandations est indispensable dans ce contexte !

Afin de résoudre cette problématique, nous devons répondre aux questions de recherche suivantes:

- Comment représenter le profil du chercheur afin de lui personnaliser les résultats de recherche?
- Quelle représentation utiliser pour les ressources scientifiques (ontologies?)
- Comment se fera la prise en compte des informations provenant des réseaux sociaux?
- Quelle architecture technique adopter pour ce système de recommandation?

III. TRAVAUX REALISES

Le travail de recherche a été réalisé en deux étapes : Etude et recherche bibliographique et contribution.

A. Etude et recherche bibliographique

La première étape dans ce travail de thèse consiste à faire une étude approfondie de la littérature. Premièrement pour enrichir mes connaissances sur un ensemble de domaines, de se familiariser avec ces concepts et d'assurer l'acquisition des bases scientifiques nécessaire à la poursuite de la thèse. Deuxièmement pour définir la direction de recherche à suivre en analysant les méthodes existantes (forces et faiblesses). Dans le cadre de notre sujet de thèse, nos recherches se focalisent principalement sur les systèmes de recommandation.

Les systèmes de recommandation:

La recommandation a été identifiée comme un moyen d'aider les individus dans une communauté pour trouver les informations ou les produits qui sont les plus susceptibles d'être intéressants pour eux ou pour être adaptés à leurs besoins [12]. Pour générer les recommandations personnalisées [13] plusieurs approches sont possibles y compris les méthodes traditionnelles telles que les méthodes à base de contenu, à base de filtrage collaboratif et hybrides [14] et récemment des méthodes avancées ont été développées, telles que les méthodes : à base de connaissances, à base de réseaux sociaux, de la confiance, sensibilité au contexte, et enfin des approches de recommandation du groupe. [15]

Les systèmes de recommandation d'articles scientifiques ont suscité beaucoup d'intérêts dans la communauté des chercheurs. Parmi les auteurs les plus productifs : C. Lee Giles qui a introduit le premier système dans le cadre du projet CiteSeer [16].

Depuis, plusieurs études, expériences, prototypes et applications ont été réalisés pour aborder divers aspects et problématiques liées à ce sujet. [17], [18], [19], [20], [21], [22],[23],[24]. Ces travaux utilisent différentes approches à savoir: Filtrage démographique, basé sur le contenu, collaboratif, Cooccurrence, à base de graphe, pertinence globale, hybride. L'approche la plus utilisée est celle basée sur le contenu (plus de 50% des travaux publiés).

D'autres travaux de recherches s'orientent vers l'intégration de l'utilisateur comme composant indispensable du modèle global de recommandation [25], [26]. En effet, la modélisation de l'utilisateur est un élément central dans ce processus pour fournir des recommandations utiles, adaptées à ses besoins précis, son contexte et ses préférences.

Une autre tendance récente dans la conception des systèmes de recommandation est l'introduction des informations issues des réseaux sociaux [28] qui représentent une véritable mine d'information. En effet, grâce aux réseaux sociaux, les individus communiquent, s'entraident, s'informent, se documentent et se font connaître. Les chercheurs et les institutions de recherche utilisent aussi de plus en plus ces réseaux sociaux, car ils leur fournissent une page clé en main qui résume leur parcours, dans le but d'accroître leurs visibilité et promouvoir les travaux de recherches [29]. En d'autres termes, le réseau social dans un système de recommandation joue au moins deux rôles importants [24] : aider les individus à améliorer leur expérience grâce à l'interaction avec leurs amis, aider également le système de recommandation à diffuser des recommandations plus précises.

À travers nos recherches, [30], [31], [32] nous avons pu constater aussi que le Machine Learning (ML) s'est imposé ces dernières années et a connu un succès remarquable dans le domaine des systèmes de recommandation.

B. Contribution

Le défi d'améliorer les pratiques et les méthodes utilisées pour rendre les systèmes plus précis, performants et adaptés aux chercheurs qui investissent dans le domaine de la recherche et consacrent la majeure partie de leur temps de travail à l'expérimentation et au progrès de leur discipline nous motive donc à proposer, dans la deuxième étape dans ce travail, une approche de recommandation basée sur l'apprentissage automatique (ML) et qui va nous permettre d'atteindre l'objectif de cette thèse en prenant en compte les caractéristiques des articles scientifiques et les profils des chercheurs. La figure ci-dessous représente le premier prototype.

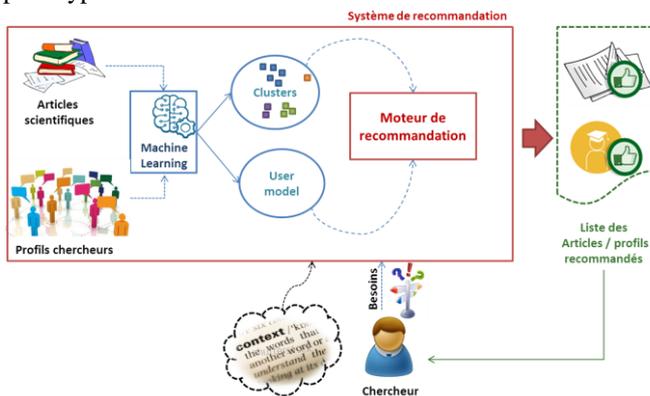


Figure 1: Prototype de notre approche de recommandation

IV. PERSPECTIVES

Nous planifions pour l'année prochaine la finalisation de notre approche de recommandation qui fera le sujet d'une communication.

A long terme les travaux qui nous restent sont:

- La validation de l'approche proposée en faisant des tests avec des données simulées ou réelles ;
- La rédaction et soumission de la publication ;
- La rédaction du manuscrit de thèse et la préparation de la soutenance.

V. CONCLUSION

À travers ce rapport, nous avons présenté l'état d'avancement de nos travaux de thèse. Après un rappel du sujet et de la problématique traitée, nous avons présenté brièvement les travaux réalisés ainsi que les perspectives fixées pour les années qui suivent.

Références

- [1] L. BERKANI, et O. NOUALI Semantic, "Collaborative Filtering for Learning Objects Recommendation". In : KDIR/KMIS. 2013. p. 52-63.
- [2] L. BERKANI, A. CHIKH et O. NOUALI. "Recommandation personnalisée des ressources dans une communauté de pratique de e-learning. Une approche à base de filtrage hybride." In : INFORSID. 2013. p. 131-138.
- [3] C. BLANC, "Pratiques et usages numériques", H2PTM'13, Imad Saleh, Manuel Zacklad, Sylvie Leleu-Merviel, Yves Jeanneret, Luc Massou, Ioan Roxin, François Soulages, Nasreddine Bouhaï (Coord. par). Hermes Science publications, Lavoisier, 2013. Revue française des sciences de l'information et de la communication, 2014, no 5.
- [4] S.B.A.L GRANGET, "L'infobésité, un phénomène au détriment de la qualité informationnelle. Pratiques et usages numériques": H2PTM'13, 2013, p. 13.
- [5] Facebook, réseau social générique. Url: <https://fr-fr.facebook.com/> [Sept. 1, 2017]
- [6] Academia.edu, réseau social scientifique. Url : <https://www.academia.edu/> [Sept. 1, 2017]
- [7] Researchgate.net, réseau social scientifique. Url : <https://www.researchgate.net> [Sept. 1, 2017]
- [8] Google.com, moteur de recherche générique. Url : <https://www.google.com> [Sept. 1, 2017]
- [9] Bing.com, moteur de recherche générique. Url : <https://www.bing.com/> [Sept. 1, 2017]
- [10] citeseer, moteur de recherche scientifique Url: <https://citeseer.ist.psu.edu/> [Sept. 1, 2017]
- [11] Scholar.google, Service de recherche d'articles scientifique. Url : <https://scholar.google.fr/> [Sept. 1, 2017]
- [12] N. MANOUSELIS, H. DRACHSLER, K. VERBERT et al., "TELas a Recommendation Context." In : Recommender Systems for Learning. Springer New York, 2013. p. 21-36.
- [13] J. LU, D. WU, M. MAO et al. "Recommender system application developments: a survey". Decision Support Systems, 2015, vol. 74, p. 12-32.
- [14] R. Burke, "systèmes de recommandation hybrides: enquête et expériences, modélisation de l'utilisateur et l'interaction de l'utilisateur Adapté", 2002, p. 331-370.
- [15] J. BEEL et S. LANGER, "A comparison of offline evaluations, online evaluations, and user studies in the context of research paper recommender systems". In : International Conference on Theory and Practice of Digital Libraries. Springer International Publishing, 2015. p. 153-168.
- [16] Giles, C. Lee, Kurt D. Bollacker, and Steve Lawrence. "CiteSeer: An automatic citation indexing system." Proceedings of the third ACM conference on Digital libraries. ACM, 1998.
- [17] H. LIU, Z. YANG, I. LEE et al., "Car: Incorporating filtered citation relations for scientific article recommendation". In : 2015 IEEE International Conference on Smart City/SocialCom/SustainCom. IEEE, 2015. p. 513-518.

- [18] K. Sugiyama and M.-Y. Kan, "Scholarly paper recommendation via user's recent research interests," in Proc. 10th Annu. Joint Conf. Digit. Libraries, 2010, pp. 29–38.
- [19] T. Strohman, W. B. Croft, and D. Jensen, "Recommending citations for academic papers," in Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2007, pp. 705–706.
- [20] F. Meng, D. Gao, W. Li, X. Sun, and Y. Hou, "A unified graph model for personalized query-oriented reference paper recommendation," in Proc. 22nd ACM Int. Conf. Inform. Knowl. Manage., 2013, pp. 1509–1512.
- [21] X. Tang, X. Wan, and X. Zhang, "Cross-language context-aware citation recommendation in scientific articles," in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2014, pp. 817–826.
- [22] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles, "A neural probabilistic model for context based citation recommendation," in Proc. 29th Conf. Artif. Intell., 2015, pp. 2404–2410.
- [23] Y. Jiang, A. Jia, Y. Feng, and D. Zhao, "Recommending academic papers via users' reading purposes," in Proc. 6th ACM Conf. Recommender Syst., 2012, pp. 241–244.
- [24] G. Tian and L. Jing, "Recommending scientific articles using birelational graph-based iterative RWR," in Proc. 7th ACM Conf. Recommender Syst., 2013, pp. 399–402.
- [25] Sahijwani, Harshita, and Sourish Dasgupta. "User Profile Based Research Paper Recommendation." arXiv preprint arXiv:1704.07757 (2017).
- [26] J. Beel. "Towards effective research-paper recommender systems and user modeling based on mind maps." *arXiv preprint:1703.09109* (2017).
- [27] Beel, Joeran, Stefan Langer, and Bela Gipp. "TF-IDuF: A Novel Term-Weighting Scheme for User Modeling based on Users' Personal Document Collections." Proceedings of the 12th iConference. To appear in March 2017.
- [28] F. M. F. WONG, Z. LIU et M. CHIANG. "On the efficiency of social recommender networks". In : 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015. p. 2317-2325.
- [29] E. BAKSHY, I. ROSENN, C. MARLOW et al., "The role of social networks in information diffusion". In : Proceedings of the 21st international conference on World Wide Web. ACM, 2012. p. 519-52
- [30] R. Herbrich, "Machine Learning at Amazon". In : Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. ACM, 2017. p. 535-535.
- [31] J. Wei, J. He, K. Chen, *et al.* "Collaborative filtering and deep learning based recommendation system for cold start items". *Expert Systems with Applications*, 2017, vol. 69, p. 29-39.
- [32] M. Nilashi, , K. B. Fard, M. Rahmani, "A Recommender System for Tourism Industry Using Cluster Ensemble and Prediction Machine Learning Techniques." *Computers & Industrial Engineering* (2017).

Approche intégrée pour l’ordonnancement de la production et la maintenance prédictive basée sur le pronostic industriel

Ladj Asma

Directeurs de thèse : Benbouzid-Si Tayeb Fatima, Varnier Christophe

Équipe : Optimisation

Abstract—Dans le cadre de l’évaluation de l’état d’avancement de notre thèse de doctorat LMD qui traite le problème d’intégration de la maintenance prédictive basée sur le pronostic dans l’ordonnancement des systèmes de production, nous allons présenter dans ce papier les travaux déjà achevés, ceux en cours de réalisation ainsi que les perspectives.

I. INTRODUCTION

Une des préoccupations majeures du monde industriel, est d’avoir un système opératoire performant et une gestion de production fiable. Ces objectifs sont conditionnés par la maîtrise de la sûreté de fonctionnement et l’augmentation de la disponibilité de l’outil industriel (machines) à moindre coût. Cela est garanti par l’adoption d’une bonne stratégie de maintenance.

Etablir séparément un calendrier de la production et un autre de la maintenance engendre des situations conflictuelles qui doivent être évitées en programmant d’une manière conjointe les tâches des deux services. Cela est connu dans la littérature sous le nom de “*production scheduling with availability constraints*” ou “*joint production and maintenance scheduling*” [1]. Traditionnellement, le service de maintenance avait pour principal objectif la réparation des équipements qui sont déjà tombés en panne. Ce type de maintenance, dite corrective, a ensuite été peu à peu complété par une approche dite préventive qui a pour objectif la réduction de la probabilité de défaillance et ainsi réduire les coûts engendrés par l’arrêt accidentel du système de production caractérisé par une période de remise en service importante.

Les travaux recensés dans la littérature tenant compte des opérations de maintenance préventive dans l’ordonnancement de la production peuvent être regroupés en deux catégories. Le cas déterministe où les interventions de maintenance dite *préventive systématique* sont périodiques et fixées à l’avance [2]. Pour le cas non déterministe, les interventions de maintenance dite *prédictive* sont programmées selon la prédiction de l’évolution de la dégradation de l’équipement [3].

II. PROBLÉMATIQUE

Une maintenance préventive peut être envisagée pour anticiper les pannes des équipements et les corriger avant de générer de trop importants dégâts qui pourraient provoquer un arrêt imprévu du système. Elle repose uniquement sur des analyses de fiabilité pour établir un calendrier des interventions fixé au préalable qui ne tient pas compte des sollicitations influençant réellement les équipements du système tout au

long de son fonctionnement. En effet, des sollicitations anormales ou imprévues peuvent accélérer la dégradation des équipements.

Pour palier à ce problème, les politiques modernes tendent vers la mise en œuvre de maintenance dite prédictive. Ces politiques ne fixent pas à l’avance la date exacte des opérations de maintenance mais sont conditionnées par le suivi en temps réel de l’état de santé de l’équipement à maintenir pour pouvoir prédire son état futur [3].

Dans ce cadre, le domaine du PHM (Prognostics and Health management) est très actif en matière de recherche pour proposer des techniques permettant notamment de prédire l’état futur de fonctionnement de l’équipement. Ainsi, il apparaît très intéressant de combiner la planification de la production dans un atelier avec la mise en place de telle politique de maintenance. De nouveaux problèmes d’ordonnancement sont alors à aborder. Ils doivent tenir compte de l’état de santé courant et des sollicitations futures des équipements.

Nous avons décidé de nous concentrer sur la prise de décision post-pronostic. Pour cela, nous supposons dans nos travaux que le système est doté d’un module PHM qui fournit des résultats les plus représentatifs de son état de santé. Notre rôle est alors d’interpréter et d’intégrer les données de sortie du pronostic dans l’ordonnancement de la production.

Il existe relativement peu de travaux intégrant la maintenance prédictive. Pour la majorité de ces travaux, le problème est modélisé sous la forme d’un programme linéaire en nombres entiers (PLNE) pour trouver le meilleur ordonnancement optimisant un certain critère lié à l’aspect de la production et de la maintenance. Nous avons opté alors pour une approche métaheuristique pour résoudre des instances du problème de grandes tailles.

III. TRAVAUX RÉALISÉS

A. Stage scientifique

Dans le cadre du Programme National Exceptionnel PNE 2016/2017, j’ai effectué un stage au sein de l’équipe PHM (Prognostic Health Management) sous la direction de mon co-encadreur Dr. Christophe Varnier.

Les objectifs de ce séjour scientifique étaient:

- Le positionnement de nos travaux de recherche par rapport à ce qui est fait dans le domaine de la maintenance prédictive.
- La finalisation des approches globales proposées.
- La publication des contributions.

B. Conférences

Durant cette année, nous avons pu participer à une conférence internationale KES'2017 (21st International Conference on Knowledge-Based and Intelligent Information & Engineering Systems). Le papier présenté intitulé: "A Hybrid of Variable Neighbor Search and Fuzzy Logic for the permutation flowshop scheduling problem with predictive maintenance" propose une méthode basée sur une recherche locale à voisinages multiples pour la résolution du problème d'ordonnancement intégré de la production et la maintenance prédictive dans un environnement incertain [4].

C. Publications

- Nous avons rédigé notre première contribution qui traite le problème d'ordonnancement intégré de la production et la maintenance prédictive au sein d'un atelier mono-machine. Ce papier a été soumis au journal IJPR (International Journal of Production Research). Malgré la réponse défavorable, nous avons profité des différentes remarques des reviewers pour améliorer notre papier.

- L'article précédemment rédigé a été adapté, corrigé et resoumis au journal IJPHM (International Journal of Prognostic and Health Management). Cet article a été accepté pour publication.

- Nous avons rédigé notre deuxième contribution qui traite le problème d'ordonnancement intégré de la production et la maintenance prédictive au sein d'un atelier multi-machines de type flowshop de permutation. Ce papier a été soumis au journal IJPR (International Journal of Production Research). La décision initiale n'est pas encore prise.

IV. PERSPECTIVE

- Validation du plan global de la thèse et mise à jour de la partie état de l'art de la thèse.

- Rédaction des chapitres de la thèse.

- Préparation d'un papier de conférence pour l'algorithme génétique proposé pour résoudre le problème d'ordonnancement intégré de la production et la maintenance prédictive au sein d'un atelier multi-machines de type flowshop de permutation.

V. CONCLUSION

Le présent document résume les principaux travaux réalisés au cours de notre quatrième année de thèse. Nous avons pu valider les différentes approches de résolutions du problème traité que ce soit dans le cas à une ou plusieurs machines. La phase actuelle a pour objectif principal de rédiger la thèse.

REFERENCES

- [1] S. Wang and M. Liu, *branch and bound algorithm for single-machine production scheduling integrated with preventive maintenance planning*, International Journal of Production Research, vol. 51, pp. 847-868, Feb. 2013.
- [2] Y. Ma, C. Chu and C. Zuo, *A survey of scheduling with deterministic machine availability constraints*, Computers Industrial Engineering, 2(58), pp 199-211, 2010.

- [3] W. Wang and M. Pecht, *Economic analysis of canary-based prognostics and health management*, IEEE Trans. Ind. Electron., vol. 58, no. 7, pp. 3077-3089, Jul. 2011.
- [4] A. Ladj, F. Benbouzid Si Tayeb, C. Varnier, A. A. Dridi and N. Selmane. "A Hybrid of Variable Neighbor Search and Fuzzy Logic for the permutation flowshop scheduling problem with predictive maintenance", Procedia Computer Science, vol.112, pp. 663-672, September. 2017.

Publication de Données Massives Préservant l'Information Sensible

Safia Bourahla

Directeur de thèse : M. Yacine CHALLAL

Équipe : SURES

Résumé : la prolifération des réseaux sociaux (SNs) a permis de créer une grande quantité de données sur les utilisateurs et les interactions entre eux. Ces données contiennent des informations privées sur les utilisateurs donc une tâche d'anonymisation est requise avant de les publier. La majorité des travaux d'anonymisation proposent des techniques qui permettent de publier une seule instance de SN sans violer la confidentialité des utilisateurs. Cependant la publication des versions séquentielles d'un SN est utile pour analyser son évolution. Dans ce résumé nous présentons un état de l'art sur les techniques d'anonymisation proposées. Nous concluons par donner des perspectives sur les problèmes ouverts à étudier dans les futurs travaux de recherche.

I. INTRODUCTION :

En créant des profils qui contiennent des informations personnelles, les réseaux sociaux permettent aux utilisateurs de créer et de rejoindre des groupes ayant des intérêts différents : politiques, économiques et géographiques. Les informations des utilisateurs et des groupes ainsi que l'affiliation des utilisateurs aux groupes sont considérées comme des données riches pour les data miners. Donc la publication des versions séquentielles de ces données répond aux exigences des data miners et leur permet de suivre l'évolution de SN. Cependant cela viole la confidentialité des utilisateurs car un adversaire peut inférer par exemple l'affiliation des utilisateurs aux groupes, qui est considérée comme information privée. Pour résoudre ce problème, les chercheurs ont proposé plusieurs techniques pour anonymiser les données avant leur publication afin que la confidentialité des utilisateurs soit préservée et les besoins des data miners soient satisfaits. Un inconvénient de ces techniques est qu'elles n'ont considéré que le cas de la publication d'une seule version de SN. Cependant, pour mieux répondre aux besoins des data miners, il est recommandé de publier des versions séquentielles anonymes pour le même réseau social afin de refléter son évolution dans le temps. Il est clair que l'application directe de ces techniques à la publication de versions séquentielles, en anonymisant chaque version indépendamment, entraîne des violations de la confidentialité car en comparant les différentes versions publiées, un attaquant peut inférer des informations privées et violer la confidentialité de l'utilisateur. Ainsi, le cas de la publication des versions séquentielles est plus difficile à assurer, car la solution proposée devrait prendre en compte les versions précédentes publiées lorsqu'elle essaie d'anonymiser les données actuelles.

II. PROBLÉMATIQUE

Le problème de confidentialité dans les SNs peut être formulé comme suit : étant donné un graphe (représentant un SN) G , le

but est de construire un graphe anonyme \bar{G} à partir du G tel que \bar{G} permet d'effectuer des analyses utiles tout en préservant la confidentialité des utilisateurs.

III. TRAVAUX RÉALISÉS

Les solutions proposées peuvent être groupées en cinq catégories : k -anonymization, Random graph editing, Clustering-based, Probabilistic privacy preservation and Differential privacy approaches. Jusqu'à présent, la plupart des travaux existants de ces catégories de solutions traitent le problème de la confidentialité dans le contexte de la publication d'une seule version des données des réseaux sociaux. Nous donnons des exemples de ces travaux dans chaque catégorie:

1) K -anonymization approaches :

Cette catégorie garantit qu'un adversaire ayant un certain background A ne peut pas identifier un individu avec une probabilité $> 1/k$. Selon le background de l'adversaire plusieurs principes ont été proposés :

- k -degree anonymity [1] : Son but est d'empêcher un adversaire, qui a comme background le degré d'un nœud v , de l'identifier dans le graphe publié. Donc il vise à atténuer les risques de divulgation de l'identité. Pour empêcher aussi les risques de divulgation du contenu ce principe a été étendu au principe k -degree- l -diversity en [2],
- k -neighborhood anonymization [3]: Son but est d'empêcher les risques de divulgation de l'identité, en considérant un adversaire ayant le voisinage d'un nœud v comme background. Dans [4] le principe de k -neighborhood l -diversity a été proposé, ce principe est une extension de principe k -neighborhood, pour empêcher les risques de divulgations du contenu.

2) Random graph editing approaches :

Les méthodes d'anonymisation de cette catégorie modifient le graphe par l'ajout, ou la suppression ou la permutation aléatoire des arêtes. Parmi les travaux proposés dans cette catégorie on trouve [5] le but de ce travail est d'atténuer les risques de divulgations de l'identité. Les solutions de cette catégorie souffrent de deux inconvénients ; elles ne satisfont pas le k -anonymat, et elles préservent la confidentialité d'un sous ensemble d'utilisateurs puisqu'elles se basent sur la modification aléatoire des arêtes.

3) Clustering-based approaches :

Dans cette catégorie au lieu de modifier le graphe, on le partitionne en des clusters. Tout d'abord on crée n clusters tel que chaque cluster contient au moins k nœuds similaires, ensuite les nœuds de chaque cluster sont modifiés pour devenir indiscernables et les arrêtes sont généralisées. Parmi les travaux proposés on a : [6] a proposé une solution pour empêcher les risques de divulgation de l'identité pour un graphe non étiqueté. Dans [7] le cas des nœuds étiquetés a été considéré, donc le but est de concevoir une solution pour les risques de divulgation du contenu. Un problème ouvert des méthodes basé sur le clustering est comment déterminer le nombre optimal de cluster, en plus ces méthodes ne publient que des clusters et des arrêtes généralisés ce qui dégrade l'utilité de données publiées.

4) Probabilistic privacy preservation approaches :

Le but de cette catégorie est de convertir un graphe déterministe en un graphe incertain par associer des probabilités aux arrêtes (la probabilité représente le niveau d'incertitude de la présence de l'arrête dans le graphe original). Parmi les travaux on cite [8] qui a proposé le principe k-obfuscation qui garantit qu'un adversaire ne peut pas déduire l'identité d'un nœud dans avec une probabilité $> \log_2 k$ et [9] qui a proposé le principe (k, ϵ) -obfuscation qui exige que le graphe soit k -obfuscates pour au moins $(1 - \epsilon)n$ nœuds.

5) Differential privacy approaches :

Initialement proposé pour les données structurées, elle garantit que en examinant les réponses des requêtes, un adversaire ne peut pas déduire la présence d'un individu dans la base de données. Dans les réseaux sociaux on dit qu'un algorithme A est ϵ -differentially private s'il satisfait l'équation suivante :

$$\Pr(A(G) = s) \leq e^\epsilon \Pr(A(\bar{G}) = s)$$

Où G et \bar{G} sont deux graphes voisins, selon la notion de graphes voisins on distingue deux niveaux :

- Node level differential privacy [10] : on considère deux graphes voisins s'ils diffèrent par un seul nœud et ses arrêtes adjacentes, son but est d'empêcher les risques de divulgation de l'identité,
- Edge level differential privacy [11]: on considère deux graphes voisins s'ils diffèrent par une seule arrête, son but est d'empêcher les risques de divulgation des liens.

Il y a peu de travaux concernant les solutions qui permettent de publier des versions séquentielles des réseaux sociaux tout en préservant la confidentialité. Dans [13] Wang et al. étudient le problème de la publication en temps réel de données spatio-temporelles dans les réseaux sociaux. Ils proposent la solution RescueDP " D "REal-time Spatio-

temporal Crowd-soUrcEd Data Publishing with Differential Privacy" pour protéger la trace de mobilité des utilisateurs. Dans [12] Bhagat et al. considèrent le problème de la confidentialité dans les réseaux sociaux modélisés par un graphe bipartite; Ils cachent le mappage entre le nœud et l'entité correspondante en divisant l'ensemble des nœuds en groupes de taille k . Ils fournissent des méthodes qui utilisent des algorithmes de prédiction de lien pour modéliser l'évolution du réseau social et prédire les futures structures. La prédiction permet de choisir une anonymisation qui devrait rester sûre et utile pour les versions ultérieures. Une limite de leur solution est qu'il fournit une méthode pour anonymiser un réseau social dynamique lorsque de nouveaux nœuds et arrêtes sont ajoutés au réseau social publié et ne considèrent pas le cas de la suppression, ce qui est fréquent dans les SNS.

IV. CONCLUSION ET PERSPECTIVES:

Parmi les problèmes ouverts à étudier dans les futurs travaux de recherche on cite :

- Les différents types de données qui existent dans les réseaux sociaux comme les hyper média,
- La scalabilité et le dynamisme d'un réseau social,
- Les publications multiples,
- Proposer une mesure unifiée pour quantifier la confidentialité obtenue,
- La variation et le changement de background de l'adversaire.

V. RÉFÉRENCE :

- [1] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," In Proceedings of the SIGMOD, pp. 93–106, 2008.
- [2] M. Yuan, L. Chen, P.S. Yu, and T. Yu, "Protecting Sensitive Labels in Social Network Data Anonymization," *IEEE Transactions on Knowledge and Data Engineering*, Volume 25, Issue 3, pp. 633–647, 2013.
- [3] B. Zhou and J. Pei, "Preserving privacy in social networks against neighborhood attacks," *ICDE2008* IEEE Press, pp. 506–515, 2008.
- [4] B. Zhou and J. Pei, "The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighbourhood attacks," *Knowledge and Information Systems*, Volume 28, pp. 47–77, 2011.
- [5] X. Ying and X. Wu, "Randomizing Social Networks: a Spectrum Preserving Approach," in *SDM*, pp. 739–750, 2008.
- [6] M. Hay, G. Miklau, D. Jensen, D. F. Towsley, and C. Li, "Re-sisting structural re-identification in anonymized social networks," *Very Large Database Journal*, Volume 19, No 6, pp. 797–823, 2010.
- [7] T. Tassa and D. Cohen, "Anonymization of centralized and distributed social networks by sequential clustering," *IEEE Transactions on Knowledge and Data Engineering*, Volume 25, No 2, 311–324, 2012.
- [8] Bonchi, F., Gionis, A., & Tassa, T. (2011). Identity Obfuscation in Graphs Through the Information Theoretic Lens. *ICDE Conference*, 924-935.
- [9] P. Boldi, F. Bonchi, A. Gionis, and T. Tassa, "Injecting uncertainty in graphs for identity obfuscation," In *Proceedings of VLDB Endowment*, Volume 5, No 11, 1376–1387, 2012.
- [10] C. Borgs, J. T. Chayes and A. Smith, Private Graphon Estimation for Sparse Graphs, arXiv:1506.06162 [math.ST], April 2015.
- [11] J. Blocki, A. Blum, A. Datta, O. Sheffet, Differentially Private Data Analysis of Social Networks via Restricted Sensitivity, Proceedings of the 4th conference on Innovations in Theoretical Computer Science, 2013.
- [12] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava. "Prediction Promotes Privacy In Dynamic Social Networks", 2010.
- [13] Q.Wang, Y. Zhang, X.Lu, Z. Wang, Z. Qin, and K.Ren. "Real-time and Spatio-temporal Crowd-sourced Social Network Data Publishing with Differential Privacy", 2016.

De la coopération des productions collectives dans un contexte multi-agent: un cas d'étude de surveillance de feux de forêts

KHENIFAR Afra

Directeurs de thèse : KOUDIL Mouloud, OCCELLO Michel, JAMONT Jean Paul

Équipe : WSN

Abstract—L'objectif de ce papier est de résumer l'état d'avancement du travail de recherche dans le cadre de la thèse intitulée "De la coopération des collectifs multi-agents par l'utilisation des productions collectives" pour l'année universitaire précédente.

I. INTRODUCTION

Notre objectif pendant plusieurs années de la thèse était de bien cerner la problématique. En effet, plusieurs visions, parfois contradictoires, coexistent dans le domaine des Systèmes Multi-Agents (SMA) et précisément concernant leurs phénomènes collectifs (c.à.d., phénomènes émergents). Le temps consacré pour bien se positionner nous a permis de définir clairement les préconisations, les hypothèses et les contraintes de la thèse qui nous ont aidé à dimensionner la contribution. Cette dernière était communiquée et validée en deux conférences [1, 2]. Pendant cette année, la contribution qui est sous forme d'un modèle de coopération était détaillée dans l'article accepté à la Revue d'Intelligence Artificielle [3]. Nous avons également travaillé pendant cette année sur un article qui était soumis à la revue Future Generation Computer Systems¹. Les reviewers ont jugé la contribution très intéressante, et ont demandé de retravailler la forme qui était la principale cause derrière le refus de l'article.

Le reste du résumé de l'état d'avancement est organisé comme suit. Nous présenterons la problématique de la thèse en section II. Les travaux réalisés seront donnés en section III. Nous concluons et expliquerons le travail qui reste à réaliser dans le cadre de la thèse en section IV.

II. PROBLÉMATIQUE

De notre point de vue, une production collective est le phénomène généré collectivement par un système. Une production collective est observée au niveau externe du système qui la produit. Nous appelons un système qui génère collectivement une production collective, un collectif. Dans ce travail (comme beaucoup d'autres travaux de SMA), nous supposons que la production collective est inscrite nulle part dans le collectif qui la génère. Ainsi, aucun agent ne possède une connaissance globale de son existence. Par ailleurs, Nous considérons une production comme un objet exploitable du collectif qui la produit, aussi comme un savoir faire de tout le collectif, d'où l'intérêt de faire

coopérer des productions collectives de différents collectifs. La production collective d'un collectif étant inscrite nulle part, comment permettre la coopération par l'utilisation de productions collectives? Comment opérationnaliser cette forme de coopération?

Dans ce travail nous visons des collectifs *préexistants* à faire coopérer. La contrainte que nous nous posons est donc de les doter de la capacité de faire coopérer leurs productions collectives tout en préservant leurs fonctionnements nominaux: quelle solution proposer pour permettre de respecter cette contrainte?

III. TRAVAUX RÉALISÉS

A. État de l'art

La thèse représente une intersection de plusieurs disciplines: SMA, systèmes embarqués et réseaux ad-hoc hétérogènes. Dans la discipline des SMA, nous travaillons sur plusieurs thématiques: coopération de SMA, phénomènes émergents, méta-connaissances et agents virtuels. Un état de l'art lié à chaque discipline et à chaque thématique était réalisé afin de bien positionner le travail de thèse.

B. Contribution

L'approche proposée comprend trois parties :

- Un ensemble de concepts, qui introduisent notre propre vision sur les notions de collectif, de production collective et de la coopération en utilisant les productions collectives;
- Une démarche méthodologique, qui consiste à guider le concepteur pour définir les connaissances nécessaires afin d'opérationnaliser la forme de coopération proposée. La méthodologie était proposée dans l'article accepté à la conférence RJCIA [4] et est enrichie dans l'article accepté à la revue RIA [3];
- Un modèle d'opérationnalisation qui doit pouvoir être greffé sur les collectifs préexistants à faire coopérer. Le modèle se base sur les connaissances déjà définies pour permettre la coopération des productions collectives en déployant des sondes virtuelles. Le modèle a été validé à travers les communications aux conférences JFSMA'16 et RIVF'16 [2, 1].

¹<https://www.journals.elsevier.com/future-generation-computer-systems>

C. Évaluation

Afin de valider le modèle, nous l'avons testé avec le simulateur MASH [5] qui était développé au sein de l'équipe, et qui est dédié à la simulation des systèmes embarqués (Fig 1).

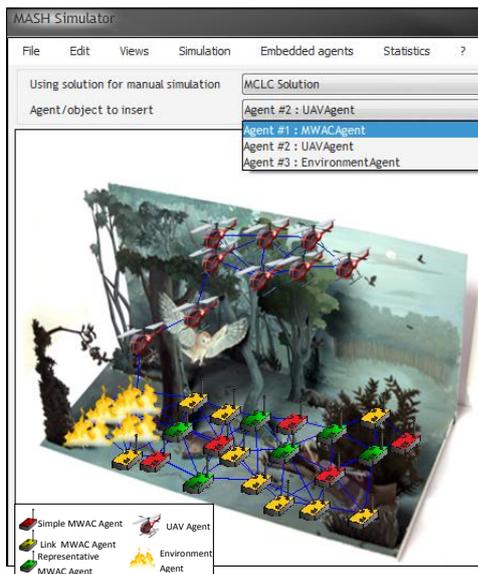


Fig. 1: Implémentation et évaluation du modèle de coopération proposé [1].

1) *Implémentation*: différentes parties de la solution étaient implémentées :

- Un réseau de drones sous le simulateur MASH;
- Une sonde virtuelle générique, de sorte qu'elle soit paramétrable pour différents cas d'étude;
- Les scénarios de l'utilisation de notre modèle pour faire coopérer un réseau de capteurs (déjà implémenté sous MASH [6]) et le réseau de drones implémenté;
- Une solution de comparaison appelée Agilla [7] qui est un middleware open source² visant à faire coopérer des réseaux hétérogènes.

2) *résultats*: grâce à la simulation, nous avons obtenu des résultats qui nous ont permis de valider l'apport du modèle en terme du temps nécessaire pour réaliser l'objectif de la coopération, qui est dans notre cas le rapprochement du feu. Cet apport a certainement un coût. Nous avons donc également évalué le coût en terme du volume moyen supplémentaire échangé par les sondes virtuelles afin de mettre en œuvre la coopération par l'utilisation de productions collectives. Les résultats obtenus pour différentes tailles du réseau de capteurs et du réseau de drones ont montré que le volume échangé est proportionnel aux tailles des deux réseaux. Pour les différentes tailles, le coût reste plus intéressant en utilisant notre modèle de coopération par rapport au modèle Agilla.

²<http://mobilab.wustl.edu/projects/agilla/>

IV. CONCLUSION ET PERSPECTIVES

Étant en phase de finalisation de la thèse, nos principales perspectives sont de terminer la rédaction de la thèse et d'avoir l'acceptation d'un article journal dans une revue internationale bien classée. Par rapport à la rédaction de la thèse, nous avons validé le plan et nous avançons dans la rédaction. Concernant l'article journal, nous rédigeons un nouvel article en prenant en considération les retours de l'article refusé par la revue Future Generation Computer Systems. Pour renforcer les chances d'acceptation de ce nouvel article journal, nous comptons implémenter notre modèle sur une plateforme réelle de capteurs et de drones.

Sur le long terme, nous visons d'instancier le fonctionnement du modèle sur un autre cas d'étude de WoT[8].

REFERENCES

- [1] A. Khenifar-Bessadi, J.-P. Jamont, M. Occello, C.-B. Ben-Yelles, and M. Koudil, "About cooperation of multiagent collective products: An approach in the context of cyber-physical systems," in *Proceedings of IEEE-RIVF*. IEEE, 2016.
- [2] A. Khenifar-Bessadi, J.-P. Jamont, M. Occello, C.-B. Ben-Yelles, and M. Koudil, "De la coopération des productions collectives dans un contexte multi-agents," in *Proceedings of Journée Francophones des Systèmes Multi-Agents (JF-SMA'16)*. Cépaduès Editions, 2016, pp. 43–52.
- [3] A. Khenifar-Bessadi, J. Jamont, M. Occello, C. Ben-Yelles, and M. Koudil, "De la coopération de collectifs multi-agents. un modèle d'utilisation de productions collectives," *Revue d'Intelligence Artificielle*, vol. 31, no. 1-2, pp. 97–132, 2017.
- [4] A. Khenifar-Bessadi, J.-P. Jamont, M. Occello, and M. Koudil, "Vers une coopération des collectifs de systèmes multi-agents hétérogènes," in *Rencontres de Jeunes Chercheurs en Intelligence Artificielle (RJCIA)*, 2015.
- [5] J.-P. Jamont, M. Occello, and E. Mendes, "Decentralized intelligent real world embedded systems: a tool to tune design and deployment," in *PAAMS*. Springer, 2013, pp. 133–144.
- [6] J.-P. Jamont, M. Occello, and A. Lagrèze, "A multiagent approach to manage communication in wireless instrumentation systems," *Measurement*, vol. 43, no. 4, pp. 489–503, 2010.
- [7] C.-L. Fok, G.-C. Roman, and C. Lu, "Agilla: A mobile agent middleware for self-adaptive wireless sensor networks," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 4, no. 3, p. 16, 2009.
- [8] A. Khenifar, J.-P. Jamont, M. Occello, C.-B. Ben-Yelles, and M. Koudil, "A recursive approach to enable the collective level interaction of the web of things applications," in *Proceedings of IWWISS 2014*. ACM, 2014, pp. 1–5.

Exploitation des réseaux sociaux pour l'amélioration des performances du BPM

Samiha El Hamali

Directrice de thèse : Latifa Mahdaoui

Équipe de Recherche : KMEELEA

Résumé - Notre travail de thèse consiste à montrer l'importance de l'utilisation des données issues des réseaux sociaux (RS) pour l'amélioration des processus métiers de l'entreprise. L'utilisation des RS peut être dans la phase de diagnostic afin de permettre une amélioration continue des processus métiers. Ces données seront la source de calcul de KPI externes qui seront comparés et corrélés avec les KPI internes. Nous appliquons notre modèle sur un cas d'utilisation dans la formation professionnelle.

I. INTRODUCTION

L'idée d'inclure les données des RS dans le cycle de vie de BPM permet aux entreprises d'être toujours à l'écoute de ses clients, avoir les avancées des concurrents, et suivre les tendances du marché. Nous proposons dans ce document la problématique de la thèse, les travaux réalisés, quelques perspectives et nous terminons par une conclusion.

II. PROBLEMATIQUE

Les réseaux sociaux (RS) contiennent des grandes masses d'information et des données non structurées qui ne sont pas encore exploitées pour améliorer les performances du BPM. La conception des applications du business process management (BPM) nécessitent la prise en compte de tous les facteurs internes et externes de l'organisation comme les feedbacks des clients et les avancées et innovations des concurrents sur le marché. L'objectif de cette thèse est l'étude de possibilités d'intégration des données issues des RS dans le cycle de vie du BPM.

Actuellement, les organisations utilisent leurs données internes (qui représentent le résultat de la phase d'exécution) pour

mesurer leurs objectifs (définis dans la phase de diagnostic) par l'utilisation des indicateurs de clé de performance ou Key Performance Indicator (KPI). Les KPI mesurés en internes sont soumis à des facteurs internes à l'entreprises et seront sujet à des interprétations faites par des acteurs internes ce qui va augmenter la subjectivité dans la prise de décisions. Nous proposons l'utilisation de sources externes basées sur les données issues des intervenants qui existent dans l'environnement de l'entreprises comme les clients, les fournisseurs et les concurrents qui sont présent sur les RS.

III. TRAVAUX REALISEES

A. Etude des notions du BPM

1. Définition de processus :

[1] définit un processus comme suit « un processus d'entreprise est un ensemble d'activités, entreprises dans un objectif déterminé. La responsabilité d'exécution de tout ou partie des activités par un acteur correspond à un rôle. Le déroulement du processus utilise des ressources et peut être conditionné par des évènements d'origine interne ou externe. L'agencement des activités correspond à la structure du processus ».

Les processus métiers sont les processus représentatifs des activités de l'entreprise indépendamment des moyens humains et techniques. Ces processus interfèrent de manière transversale dans le système d'information et peuvent même traverser les frontières organisationnelles internes de l'entreprise. Les processus métiers traversent même les limites de l'entreprise pour collaborer avec les partenaires comme les

fournisseurs et les distributeurs. La figure suivante montre un méta-modèle de processus proposé par [2].

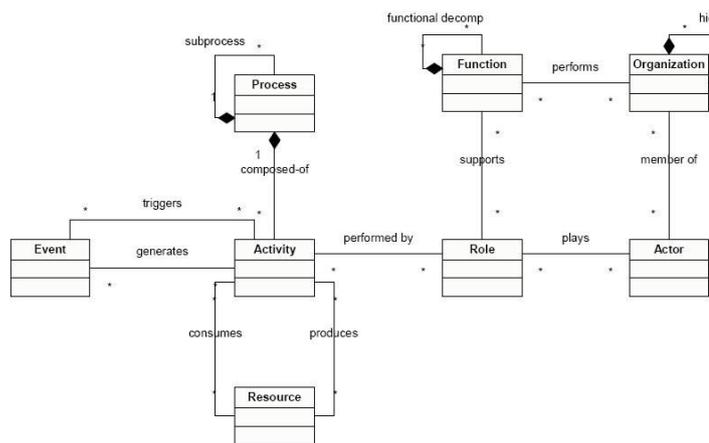


Figure 1 : un méta-modèle de processus [2]

Les activités d'un processus métier sont exécutées par des acteurs jouant des rôles particuliers, consommant quelques ressources et produisant d'autres. Les activités peuvent être déclenchées par des événements et peuvent à leur tour produire des événements. Les activités d'un processus peuvent être liées par des dépendances de ressource (dépendances de producteur-consommateur) ou des dépendances de commande (une activité déclenchant une autre). Les acteurs opèrent à l'intérieur des frontières des organisations qui exécutent des fonctions métiers spécifiques. Les rôles peuvent soutenir des fonctions. [2].

Un exemple de processus peut être le processus « fabrication de produit » qui est la combinaison d'activités pour transformer la matière achetée en produit fini.

2. Définition du BPM :

Le BPM est la gestion des processus métiers de l'entreprise. [3] définit un BPM comme suit : " gestion des processus métiers en utilisant des méthodes, techniques et logiciels pour concevoir, mettre en place, contrôler et analyser les processus opérationnels impliquant des êtres humains, organisations, applications, documents et d'autres sources d'informations". Notons que cette définition exclut les processus stratégiques et peu structurés. Le BPM se base essentiellement sur des systèmes de gestion de workflow pour la gestion automatique des processus (en anglais Workflow Management System

(WFMS). La WfMC1 définit un workflow comme " l'automatisation d'une partie ou la totalité d'un processus métier durant laquelle les documents, informations, ou tâches vont passer d'un participant à un autre pour action selon un ensemble de règles". La WfMC définit un WFMS comme suit "un système qui définit, crée et gère l'exécution des workflows à travers l'utilisation de logiciels qui s'exécute sous un ou plusieurs moteurs de workflow qui est capable d'interpréter la définition de processus, interagir avec les participants du workflow et si nécessaire invoquer l'utilisation des outils et applications informatiques".

3. Typologie des processus métiers :

Selon le framework de classification des processus métiers(PCF2) d'APQC3 deux types de processus métiers existent : les processus opérationnels, et les processus de gestion et de support. Le PCF a été développé à partir de 1992, en 2008 l'APQC a travaillé avec IBM pour améliorer le framework de processus allant même jusqu'à la spécification de framework de processus pour certaines industries. Le PCF considère 12 niveaux de groupes de processus4 (dont 5 sont opérationnels et 7 sont de gestion et de support) incluant plus de 1000 processus.

B. Etude des différents cycles de vie BPM

Plusieurs cycles de vies pour le BPM ont été proposé par plusieurs chercheurs tels que : [4] et [5].

Nous avons utilisé et choisi le cycle de vie de Van Der Aalst comme référence pour l'amélioration de ce dernier.

C. Amélioration du cycle de vie BPM par l'ajout des RS

Nous avons fait la proposition d'une première architecture de système en utilisant les données issues des réseaux sociaux pour une réingénierie du cycle de vie des processus métiers classique.

- L'étape qui sera touchée par le changement est l'étape de diagnostic,

¹ WfMC : Workflow Management Coalition

² PCF : Process Classification Framework

³ APQC: American Productivity and Quality Center.

⁴ Voir plus loin dans la section 4(possibilité d'extension du cycle de vie par les RS)

- Des KPI seront calculés à travers les données des RS
- Ces KPI seront comparées à ceux issues des données interne de l'entreprise.

Nous avons réfléchi de façon générale indépendamment du métier à inclure les données du RS pour améliorer les processus métier. Si nous revenons au modèle de cycle de vie de BPM, nous allons procéder par élimination des phases suivantes : la phase de conception et d'implémentation. Par contre, nous pouvons inclure les données des RS dans la phase opérationnelle et la phase de diagnostic. Dans la phase opérationnelle nous mesurons les KPI provenant des traces d'exécutions, comme nous pouvons mesurer ces mêmes KPI à travers les données collectées des RS afin de pouvoir les comparer ou tout simplement découvrir de nouveaux KPI qui seront ajoutés à la liste des objectifs à atteindre par la phase de diagnostic. Le schéma suivant illustre cette idée.

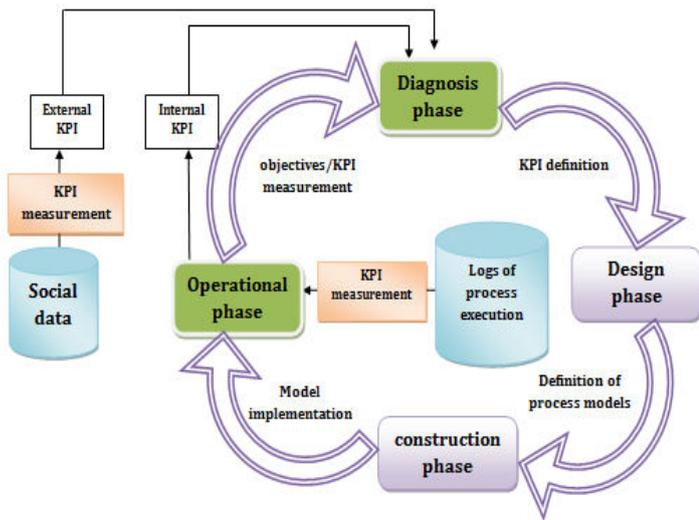


Figure 2 : ajout des données de RS au cycle de vie BPM [6]

Un exemple de KPI qui peut être étudié est la qualité de produit vendu par un concessionnaire automobile. Ce KPI peut être mesuré en interne par la mesure de nombre de réclamation de client dans le service après vente, ce même KPI peut être mesuré sur les RS par le nombre de fois ou la marque de voiture de même client a tombé en panne dans la même période.

D. Étude d'un cas concret montrant l'utilisation de RS dans le cadre de concessionnaire automobile

L'étude d'un exemple de concessionnaire automobile en Algérie à montré que les RS ont permis de changer le processus de vente et ce en analysant les commentaires des clients sur Facebook.

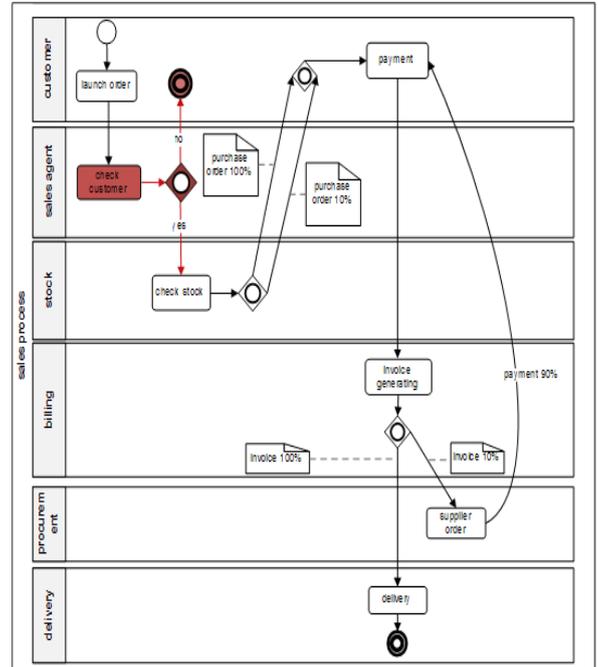


Figure 3: sales process after one iteration of BPM lifecycle using SN data [7]

E. Catégorisation des KPI

Les objectifs mesurables peuvent être calculés par les moyens des KPI. La figure suivante montre un exemple d'objectif qui peut être décomposés en sous objectifs.

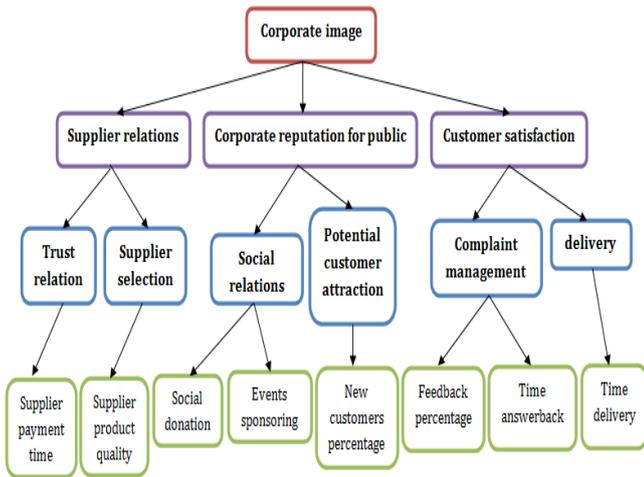


Fig.4 : exemple hiérarchie d'objectifs d'une organisation

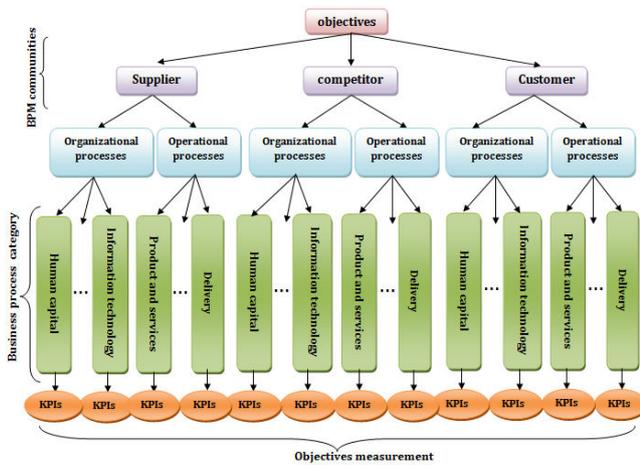


Figure 5 : niveau d'objectifs mesuré par les KPI

SN type	Implication type	Access type to SN	Data
The SN is the property of one organization	Direct	The organization defines the access type to the public: free access or controlled access.	Instance data of Forms Values or concepts extraction from free texts Analyzing of SN graph
The SN is the property of more than one organization (which form a group) and concern one domain	Indirect: when the group of organizations manages the access Direct: when each organization manages its accesses	Free and /or subscription	The same as the first case. The data is the property of the group.
Concerns one domain but not the property of any organization	Indirect	Free and /or subscription	The same as the first case.
General (like Facebook)	Limited to the list of connections	Free and /or subscription	The same as the first case.

Tab.1: different possibilities of using SN for BPM [8]

F. Proposition d'une catégorisation des différents scénarios pour l'utilisation des RS

Selon le type de processus et le propriétaire du réseau social différents scénarios peuvent exister pour utiliser les RS pour améliorer le BPM. Nous avons fait une étude synthétique pour les différents scénarios et possibilités qui peuvent être implémenté un tel système. Le tableau suivant présente un sommaire pour ces différentes situations.

G. Proposition d'un méta-modèle pour l'utilisation des RS dans un contexte BPM

Nous travaillons actuellement sur un méta-modèle pour notre solution, ce méta-modèle contient quatre vues :

1. Vue Processus:

Cette vue contient les concepts en relation avec la notion de : processus, activité, donnée interne, événement, ressource, et règle métier.

2. Vue Organisation :

Cette vue contient les concepts en relation avec la notion de : rôle, acteur humain, software et utilisateur.

3. Vue Processus :

Cette vue contient les concepts en relation avec la notion de : objectif, statut notification, KPI, KPI interne et KPI externe.

4. SN view :

Cette vue contient les concepts en relation avec la notion de : RS : communauté, concurrent, fournisseur, client, public, and data de RS.

H. Stage en France :

Stage effectué en Mai 2015 dans le cadre de la thèse à L'ENSMA.

- Connaitre l'environnement de travail du laboratoire LIAS
- Discuter les limites du sujet de recherche
- Discuter de manière générale les stratégies d'utilisation des réseaux sociaux
- Réflexion sur les différents scénarios d'exploitation des réseaux sociaux à savoir ceux qui sont à l'intérieur ou à l'extérieur de l'entreprise pour améliorer le Business Process Management
- Discuter les difficultés liées à l'utilisation des réseaux sociaux tel que les problèmes de lois et le grand volume de données.

I. Etude de cas

Une étude de cas est en cours (presque finie) pour montrer la faisabilité d'un des scénarios proposés dans le tableau précédent. Plus précisément il s'agit de montrer les différentes composante d'un système qui permet d'inclure les données extrait d'un RS général pour améliorer le processus métier dans le contexte des organismes d'éducation et de formation à but non lucratif.

J. Proposition d'un sujet de master

Un sujet de master a été proposé et soutenu en juin 2017. Le domaine visé par le travail est l'évaluation des processus dans les organismes d'éducation et de formation professionnelle (business éducation) à but non lucratif dont l'objectif est d'augmenter le chiffre d'affaire et faire le suivi de chaque formation afin d'assurer la satisfaction de client. Le cas d'utilisation proposé dans ce master a fait l'objet d'un article accepté dans la conférence AICCSA 2017[9].

Le but de travail était de montrer comment l'utilisation des réseaux sociaux peut contribuer à l'amélioration de ce genre d'entreprise. Une application a été réalisée dans ce cadre.

IV. PERSPECTIVES ET FUTURS TRAVAUX

Nous avons validé notre approche sur l'utilisation des réseaux sociaux par la solution réalisée à travers le sujet de master qui a été finalisé en juin 2017. Ce travail valide un cas d'étude qui montre la faisabilité technique de l'approche dans un contexte d'organisation à but non lucratif. Nous allons commencer la rédaction du papier journal suite à la finalisation de l'application réalisée pour les organismes de formation professionnelle. Nous allons penser aussi à la structuration de travail en vue de commencer la rédaction de la thèse.

V. CONCLUSION

Ce document montre notre problématique de recherche, les différentes étapes que nous avons suivi à savoir une étude bibliographique pour les notions autour du BPM et RS, les différents cycles de vie du BPM, la proposition d'une amélioration des scénarios existants par l'ajout de la dimension sociale, l'étude de KPI, la catégorisation des différents scénarios possibles pour l'utilisation des RS, un cas d'utilisation, la réalisation d'une application dans le cadre de la formation professionnelle. Nous continuons toujours à valider notre approche et ce par la rédaction d'un papier journal.

Références

- [1] Chantal Morley, Jean Hugues, Bernard Leblanc, and Olivier Hugues, *Processus Métiers et systèmes d'information : Evaluation, modélisation, mise en oeuvre.*: Dunod, 2005.
- [2] Hafedh Mili et al., Business process modeling languages: Sorting through the alphabet soup. *ACM Comput. Surv.* 43, 1, Article 4., novembre 2010.
- [3] Wil Van der Aalst, Arthur Ter Hofstede, and Mathias Weske, *Business Process Management: A Survey.* *BPM 2003, LNCS 2678*, pp. 1–12. Berlin Heidelberg: Springer-Verlag, 2003.
- [4] Wil Van Der Aalst and Kees Van Hee, *Workflow Management : Models, methods and systems.*, 2000.
- [5] David Hollingsworth, *The Workflow Reference Model 10 Years On.* p 295-312.: *Workflow Handbook*, 2004.
- [6] Samiha El Hamali and Latifa Mahdaoui, *Towards a Reengineering of Business Process Management Lifecycle by the Use of Social Networks.* Athens, Greece: IEEE RCIS conference, IEEE Ninth International Conference on Research Challenges in Information Science, 2015.
- [7] Samiha El Hamali & Latifa Mahdaoui : *Using Social Networks to Enhance Diagnosis in Business Process Management.* ISKO Maghreb 2015 (Tunisie)
- [8] Samiha El Hamali & Latifa Mahdaoui : *Using Facebook as external social network to enhance business processes.* The 2nd Conference on Computing Systems and Applications. CSA-2016 (Algeria)
- [9] Samiha El Hamali, Latifa Mahdaoui and Srikanth Podatharapu : *Using Social Networks to Improve Processes in Business Education: A Study Case,* AICCSA 2017 (Tunisia).

Architectures flexibles et reconfigurables pour l'implémentation de la cryptographie à base des couplages

Oussama Azzouzi

Directeurs de thèse : Mouloud Koudil et Mohamed Anane

Équipe : CoDesign

Résumé— L'objectif de nos travaux de thèse est la proposition des architectures innovantes et efficaces pour des implémentations flexibles et reconfigurables basés sur les fonctions de couplage en utilisant les circuits FPGAs de Xilinx. Durant l'année écarterée, les principales fonctions de couplage à savoir *Weil*, *Tate*, *Ate* et *Optimal Ate*, ont été implémentées et validées en coordonnées affines et jacobiennes en utilisant les courbes de *Barreto-Naehrig* (BN) sur le circuit *Virtex5* et la carte *ZedBoard*.

I. INTRODUCTION & PROBLÉMATIQUE

La sécurité de nombreux protocoles asymétriques actuels repose sur le problème du logarithme discret (DLP). L'un des premiers protocoles asymétriques est l'échange de clé de *Diffie-Hellman*. Sa sécurité repose sur la difficulté de résoudre le DLP. Actuellement, le protocole le plus célèbre est le RSA, créé par *Rivest*, *Shamir* et *Adelmann* en 1978 [1]. Sa sécurité repose sur la factorisation en nombres premiers. La taille des clés utilisées est grande (de l'ordre de 1024 bits) et doit croître au fur et à mesure que les capacités de calcul augmentent. L'apparition de la cryptographie basée sur les courbes elliptiques ECC (*Elliptic Curve Cryptography*) et HECC (*Hyper-Elliptic Curve Cryptography*) [2] a réduit sensiblement les tailles des clés des crypto systèmes à clé publique, tout en assurant le même degré de sécurité.

Les couplages sont des outils mathématiques introduits par *André Weil* en 1948 [3]. Tout d'abord, les couplages sur les courbes elliptiques ont été introduits en cryptographie dans un but de cryptanalyse. La propriété de la bilinéarité des couplages a permis de transférer le DLP sur une courbe elliptique $E(F_p)$ en un DLP sur un corps fini F_p . Ceci introduit l'attaque de MOV [4] en utilisant le couplage de *Weil* et l'attaque de *Frey* et *Rück* [5] en utilisant le couplage de *Tate*. Cependant, l'utilisation constrictive des couplages en cryptographie apparait à partir des années 2000. *Joux* [6] proposa un schéma d'échange de clé tri-partite à la *Diffie Hellman*. Le chiffrement à base d'identité de *Boneh* et *Franklin* [7] reste sans doute l'application la plus remarquable des couplages.

Un couplage est une fonction bilinéaire et non-dégénérée qui opère sur des groupes cycliques et dont le résultat est dans une extension de corps F_{p^k} .

$$e : G_1 \times G_2 \rightarrow G_3$$

De nombreux articles proposent des protocoles utilisant les couplages [8], [9] et d'autres améliorent les calculs des couplages [10], [11] et peu d'articles proposent des réalisations matérielles pour calculer les couplages [12], [13].

Afin d'exécuter les couplages d'une manière efficace, nous avons besoin d'un type particulier de courbe qu'on appelle : *pairing-friendly*. L'un des meilleurs choix en termes d'efficacité calculatoire et de sécurité est les courbes de *Barreto-Naehrig*. Ce sont des courbes elliptiques définis sur F_p avec $p \equiv 1 \pmod{12}$, de la forme $E : y^2 = x^3 + a$ avec $a \neq 0$ et dont le degré de plongement $k = 12$. Elles sont définies par un paramètre t comme suit :

$$\begin{aligned} p(t) &= 36t^4 + 36t^3 + 24t^2 + 6t + 1 \\ m(t) &= 36t^4 + 36t^3 + 18t^2 + 6t + 1 \\ t_r(t) &= 6t^2 + 1 \\ &\text{avec } t \in \mathbb{Z}. \end{aligned}$$

La construction des fonctions de couplage à savoir *Weil*, *Tate*, *Ate* et *Optimal Ate* se fait en se basant sur l'algorithme de *Miller* [14], datant de 1986, et dont les performances dépendent de l'arithmétique utilisée dans le corps de base F_p et dans ses extensions F_{p^k} . L'algorithme de *Miller* est défini comme suit :

Algorithme 1. *Miller* (P, Q, r)

Données : $r = (r_n \dots r_0)$, $P \in G_1(\subset E(F_p))$ et $Q \in G_2(\subset E(F_{p^k}))$;

Résultat : $f_{r,P}(Q) \in G_3(\subset E(F_{p^k}))$;

```

1.  $T \leftarrow P$ 
2.  $f_1 \leftarrow 1$ 
3.  $f_2 \leftarrow 1$ 
pour  $i = n - 1$  to 0 faire
    4.  $T \leftarrow [2]T$ 
    5.  $f_1 \leftarrow f_1^2 \times l_1(Q)$ ,  $l_1$  est la tangente à la courbe en  $T$ 
    6.  $f_2 \leftarrow f_2^2 \times v_1(Q)$ ,  $v_1$  est la droite verticale en  $[2]T$ 
       si  $r_i = 1$  alors
           7.  $T \leftarrow T + P$ 
           8.  $f_1 \leftarrow f_1 \times l_2(Q)$ ,  $l_2$  est la droite  $(PT)$ 
           9.  $f_2 \leftarrow f_2^2 \times v_2(Q)$ ,  $v_2$  est la verticale en  $P + T$ 
       fin
fin
retourner  $\frac{f_1}{f_2}$ 

```

La complexité de l'algorithme de *Miller* est liée à celle des fonctions l_1 , l_2 , v_1 et v_2 . Leurs complexités dépendent du système de coordonnées choisi (affine, jacobienne, ...).

Par ailleurs, les circuits FPGAs sont devenus l'option favorite pour la mise en œuvre matérielle des systèmes numériques, car ils sont disponibles et reprogrammables c'est-

à-dire, ils permettent de configurer le matériel selon les besoins, ce qui nous donne plus de souplesse pour effectuer des tests réels à des moindres coûts. En exploitant cette flexibilité, plusieurs architectures cryptographiques peuvent être développées et testées, où les parts du software et du hardware seront modulées selon les exigences de sécurité et de performances.

D'où la problématique de notre travail qui consiste à proposer des architectures innovantes et efficaces pour des implémentations software/hardware des fonctions de couplage à base des courbes elliptiques à un niveau de sécurité égal ou supérieur à 128 bits, tout en assurant un bon compromis entre le temps d'exécution et la surface matérielle occupée. Les solutions proposées seront implémentées en PSoC (Programmable System on Chip) sur circuits FPGA (Field Programmable Gate Array) de Xilinx.

II. TRAVAUX REALISES

Il existe deux grandes manières d'optimiser les calculs de couplage pour les courbes elliptiques : les optimisations arithmétiques et les optimisations mathématiques.

Les optimisations arithmétiques consistent à trouver une méthode efficace permettant le calcul de produit de deux éléments dans le corps de base F_p ou dans une extension de corps F_{p^k} . Ceci implique une étude comparative entre les différentes méthodes qui existent à savoir : la méthode naïve, *Karatsuba*, *Toom cook*, ...

Les optimisations mathématiques permettent d'apporter parfois d'énormes optimisations en temps de calcul en utilisant des techniques purement mathématique tels que : l'utilisation des twists, les corps amis, les systèmes de coordonnées, le poids de Hamming, l'exponentiation finale...

Durant cette année, les opérations de base à savoir l'addition, la soustraction, la multiplication et l'inversion ont été implémenté sur les corps F_p , F_{p^2} , F_{p^6} et $F_{p^{12}}$, en choisissant la méthode adéquate. L'addition et le doublement de point représentent les fonctions clés de l'algorithme de *Miller*, ont été implémenté et validé. Après, la méthode addition chaînée est choisie pour l'exponentiation finale. Enfin, une étude comparative entre les quatre fonctions de couplage est effectuée et ceci après avoir programmé ces fonctions en software sur les deux cartes qu'on possède.

III. PERSPECTIVES

Dans nos futurs travaux nous envisageons :

- D'appliquer le principe de multi-processing en améliorant le temps d'exécution.

- D'aller vers des architectures software/hardware en assurant un bon compromis entre le temps d'exécution et la surface occupée.

IV. CONCLUSION

Nous avons présenté brièvement notre état d'avancement au cours de cette année. Reste à finaliser les architectures proposées, avec d'éventuelles améliorations, et valoriser les travaux réalisés via les publications.

Références

- [1] R. Rivest, A. Shamir and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Communication in ACM, vol 21, pages 120-126, 1978.
- [2] Henri Cohen and Gerhard Frey, *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Chapman & Hall/CRC, Taylor & Francis group, 2006.
- [3] A. Weil, *Variétés abéliennes et courbes algébriques*, Hermann, 1948.
- [4] A. Menezes, T. Okamoto and S. Vanstone, *Reducing elliptic curve logarithms to logarithms in a finite field*, Proceedings of the twenty-third annual ACM symposium on Theory of computing STOC'91, pages 80-89, 1991.
- [5] G. Frey and H.G. Rück, *A remark concerning m-divisibility and the discrete logarithm problem in the divisor class group of groups*, In Mathematics of Computation, vol 62, pages 865-874, 1994.
- [6] A. Joux, *A one round protocol for tripartite Diffie-Hellman*, Proceedings of the 4th International Symposium on Algorithmic Number Theory, Springer-Verlag, pages 385-394, 2000.
- [7] D. Boneh and M.K. Franklin, *Identity-Based Encryption from the Weil Pairing*, Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology CRYPTO'01, Springer-Verlag, pages 213-229, 2001.
- [8] D. Boneh, B. Lynn and H. Shacham, *Short signatures from the Weil pairing*, Advances in Cryptology ASIACRYPT'01, Springer Verlag, vol 2248, pages 514-532, 2001.
- [9] D. Boneh and H. Shacham, *Group Signatures with Verifier-Local Revocation*, 11th Conference on Computer and Communications Security CCS'04, ACM, pages 168-177, 2004.
- [10] M. Scott, *Computing the Tate Pairing*, Topics in Cryptology CT-RSA'05, vol 3376, pages 293-304, 2005.
- [11] F. Vercauteren, *Optimal pairings*, IEEE Transactions on Information Theory, 56(1):455-461, January 2010.
- [12] Jean-Luc Beuchat, Jorge E. Gonzalez-Diaz, Shigeo Mitsunari, Eiji Okamoto, Francisco Roríguez-Henriquez, and Tadlnori Teruya. *High-Speed Software Implementation of the Optimal Ate Pairing over BarretoNaehrig Curves*, In Pairing-Based Cryptography-Pairing 2010, vol 6487 of Lecture Notes in Computer Science, pages 21-39, Springer, 2010.
- [13] S. Duquesne, L. Ghammam, *Memory-saving computation of the pairing final exponentiation on BN curves*, Groups Complexity Cryptology, vol 8, 2016.
- [14] V. S. Miller, *The Weil pairing, and its efficient calculation*, Journal of Cryptology, Springer-Verlag, vol 17, pages 235-261, 2004.

Cyber-Sécurité dans les Systèmes Cyber-Physiques

Mourad BENMALEK

Directeurs de thèse : Yacine CHALLAL

Équipe : SURES Research Team

Abstract—Le présent document résume les travaux de recherche déjà réalisés dans le cadre de ma thèse, présente brièvement les travaux en cours, ainsi que les perspectives d'avenir. Ces travaux consistent principalement en la proposition de solutions de sécurité dans les systèmes cyber-physiques en prenant en compte leurs caractéristiques intrinsèques et opérationnelles. Des solutions de sécurité ont été proposées et communiquées dans des conférences internationales. Une nouvelle solution a été soumise pour publication dans une revue scientifique à caractère international.

I. CONTEXTE ET PROBLÉMATIQUE

UN SYSTÈME CYBER-PHYSIQUE représente un système d'éléments de calcul qui collaborent dans le but d'assurer la coordination, la surveillance, et la commande de processus physiques. L'objectif ultime des CPS (Cyber Physical Systems) est la création d'espaces intelligents : smart health, smart cities, smart grid, smart transport, ... etc.

L'implication des CPS dans le contrôle du monde physique suscitera des questions, qui concerneront directement la sécurité des biens et des personnes. Par exemple, certaines applications peuvent être étroitement liées à des infrastructures stratégiques telles que la fourniture d'eau et d'électricité, la surveillance de ponts et bâtiments, tandis que d'autres géreront des informations liées à la vie privée des personnes comme leurs consommations énergétiques et états de santé.

L'hétérogénéité des technologies habilitantes dans les systèmes cyber-physiques et leur forte dimension humaine et socio-économique rendent leur sécurité un sujet difficile et complexe. En plus des problèmes de sécurité des technologies qui les constitueront et qui concerne en premier lieu la sécurité des données (intégrité, authenticité et confidentialité), des communications et des infrastructures réseaux, les CPS accentuent les problèmes de sécurité des personnes qui les utiliseront.

Par ailleurs, la prévention et détection d'attaques et d'intrusions dans les CPS doit faire face à de nouveaux challenges comme : le passage à l'échelle, la limitation de ressources des objets qui les constitueront, les données massives générées par de tels systèmes, ... etc. qui rendent les techniques traditionnelles inefficaces.

D'où la problématique de notre travail de recherche qui consiste en la proposition de solutions de sécurité dans les CPS en prenant en compte leurs caractéristiques intrinsèques et opérationnelles.

II. TRAVAUX RÉALISÉS

Les travaux effectués durant cette année de thèse étaient une continuité logique des travaux de recherche déjà entamés en première, deuxième, et troisième année de thèse. Dans cette section, nous présentons les différents travaux qui ont été réalisés et publiés .

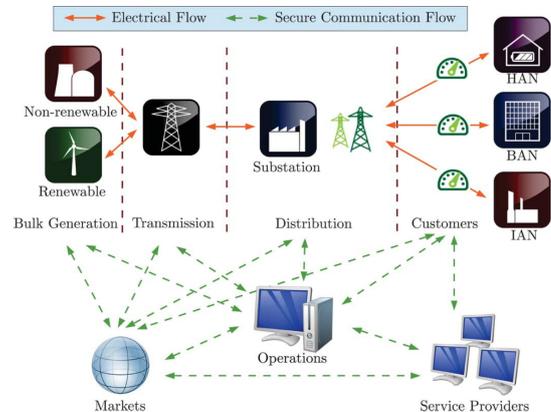


Fig. 1. Architecture d'un "Smart Grid"

A. Périmètre ciblé

Nous avons ciblé dans nos travaux de recherche de la deuxième, troisième, et quatrième année un périmètre précis qui est la gestion de clés (Key Management) dans les systèmes électriques intelligents (Smart Grids).

Le "Smart Grid" (SG) [1] est le système électrique de nouvelle génération et qui est une des réponses au défi de gestion énergétique. Comme son nom l'indique il s'agit de rendre le réseau intelligent, ou plutôt communiquant, en doublant le réseau actuel d'une infrastructure de télécommunication reliant le producteur d'électricité au consommateur (Fig. 1). Cet échange d'informations (bidirectionnel et en temps réel) entre les différents acteurs du système électrique c'est-à-dire les producteurs et l'utilisateur final permettra ainsi un pilotage fin du réseau en temps réel. De plus, les énergies renouvelables sont mieux intégrées aux réseaux intelligents et elles deviennent plus actives dans l'équilibre de l'offre et de la demande.

Un des composants fonctionnels indispensables dans toute architecture de communication sécurisée est le sous-système de gestion de clés. Ce sous-système revêt une importance particulière dans les "Smart Grid" étant donné leurs spécificités qui rendent l'assurance de cette fonction un challenge important pour lequel il est nécessaire de trouver des solutions.

L'objectif est donc (1) comment *authentifier* les dispositifs dans un SG ? (2) comment assurer un *échange sécurisé* des informations bidirectionnelles entre les différents acteurs du "Smart Grid" pour qu'un tel système fonctionne et soit sécurisé ? chacun des dispositifs du "Smart Grid" doit disposer d'un ensemble de clés secrètes. Ceci implique la *génération* des clés, leurs *distributions* de manière sécurisée aux utilisateurs, leurs *stockages* et leurs *vérifications*. Après l'étude de la structure globale du SG, des différents types de messages échangés

et des objectifs de la sécurité dans un tel système, nous pouvons définir les exigences de sécurité pour la conception d'un système de gestion de clés pour le SG comme suit :

* **La scalabilité:** qui représente un enjeu majeur pour un tel système à grande échelle consistant en des milliers voire des millions de dispositifs électriques.

* **Les exigences en délai de transmission:** de certains types de messages.

* **L'efficacité:** nous devons prendre en considération 3 aspects qui sont : (a) la capacité réduite de stockage (les compteurs intelligents par exemple qui dispose d'une RAM de 8 Ko et d'une mémoire flash de 120 Ko), (b) la capacité réduite de calcul des dispositifs, (c) "l'overhead" de communication lors de la génération, distribution et rafraichissement des clés.

* **L'ubiquité des dispositifs intelligents du SG:** et la difficulté de leur procurer une protection physique et/ou une surveillance permanente, les exposent au risque de compromission physique par des intrus. Ceci peut avoir un impact important si les intrus réussissent à récupérer les clés cryptographiques qui seraient dans la mémoire des dispositifs corrompus et à les utiliser pour mener des attaques diverses sur le réseau et les autres objets. Afin de réduire les conséquences de cette vulnérabilité, il serait nécessaire que la gestion de clés soit résiliente, en d'autres mots "tolérante à la compromission de dispositifs".

* **Le mode de transmission hybride de messages au sein du SG:** le système de gestion de clés conçu doit supporter les 3 modes de transmission à savoir l'unicast, le broadcast et le multicast. De plus, pour chaque mode nous devons préciser clairement les méthodes de génération, de distribution et de rafraichissement des clés.

B. Approche proposées et publiées

Trois approches de gestion de clés [2] [3] [4] ont été proposées et communiquées dans des conférences internationales.

Un premier papier [2] a été soumis, accepté puis communiqué oralement à *IEEE TRUSTCOM-2015* intitulé "eSKAMI: Efficient and Scalable multi-group Key management for Advanced Metering Infrastructure in Smart Grid".

Basée sur le mécanisme LKH (Logical Key Hierarchy) [5] vu ses performances en matière de stockage des clés et en matière de bande passante nécessaire pour mettre à jour les clés, une nouvelle structure de gestion de clés pour le Smart Grid (eSKAMI) a été proposée pour sécuriser les messages transmis en unicast, broadcast et multicast.

Nous avons proposé SAMI (Scalable multi-group key management for Advanced Metering Infrastructure) [4] un mécanisme de gestion de clés efficace et scalable basé sur OFT [6] qui représente une amélioration de notre premier mécanisme eSKAMI [2] en matière de bande passante. Ce deuxième papier a été soumis, accepté puis communiqué oralement à la conférence internationale *IEEE CIT-2015*.

Un troisième papier [4] a été soumis, accepté puis communiqué oralement à *IEEE WCNC-2016*. Basée sur le mécanisme OFC (One-way Function Chain) [7], cette nouvelle structure de gestion de clés pour AMIs (Advanced Metering Infrastructures) donne de meilleurs résultats par rapport au premier mécanisme proposé (eSKAMI).

C. Nouvelle approche proposée

Un système de gestion de clés de groupes à pour fonctions l'établissement de la clés de groupes par distribution ou accord et la mise à jour des clés de groupes (*rekey*) suite à un changement dans la composition des groupes ou à cause d'une compromission. En effet, pour garantir la confidentialité future et passée, il est nécessaire de mettre à jour les clés des groupes suite à l'ajout ou suppression (départ, défaillance ou expulsion) de membres. Cette mise à jour des clés peut se faire immédiatement après l'ajout ou suppression d'un membre (*individual rekeying*), ou bien après une période bien définie par le mécanisme de gestion de clés (*batch rekeying*).

Contrairement aux trois approches proposées et basées sur l'*individual rekeying*, une nouvelle approche a été proposée (VerSAMI: Versatile and Scalable key management for Smart Grid AMI systems). Basée sur un mécanisme de *batch rekeying*, son but est de limiter le nombre de membres des groupes affectés par le calcul et la redistribution des clés suite à l'ajout ou à la suppression d'un ou plusieurs membres dans une période bien définie.

Ce papier a été soumis (*revised version*) pour une possible publication dans la revue internationale *Computer Networks*.

III. PERSPECTIVES D'AVENIR

Les travaux à effectuer durant la dernière année de thèse sont une continuité logique des travaux déjà entamés. En effet, nous envisageons de finaliser deux articles que nous comptons soumettre à des revues scientifiques à caractère international: un *survey* sur les méthodes de gestion de clés pour les Smart Grid proposées dans la littérature, et une nouvelle architecture *semi-centralisée* de gestion de clés pour les Smart Grid.

La soutenance est prévue pour Juin 2018.

IV. CONCLUSION

Au vue des travaux présentés en haut, nous estimons que l'avancement des travaux de thèse est globalement positif.

REFERENCES

- [1] NIST, "NIST framework and roadmap for smart grid interoperability standards," *release 1.0, NIST Special Publication 1108*, pp. 1-145, 2010.
- [2] M. Benmalek and Y. Challal, "eSKAMI: Efficient and Scalable multi-group Key management for Advanced Metering Infrastructure in Smart Grid," in *proceedings of The 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, Helsinki, Finlande*, pp. 782-789, Aug. 2015.
- [3] M. Benmalek, Y. Challal and A. Bouabdallah, "Scalable Multi-group Key Management for Advanced Metering Infrastructure," in *2015 IEEE CIT, Liverpool*, 2015, pp. 183-190.
- [4] M. Benmalek and Y. Challal, "MK-AMI: Efficient multi-group key management scheme for secure communications in AMI systems," in *2016 IEEE WCNC, Doha*, 2016, pp. 1-6.
- [5] C. K. Wong, M. Gouda, and S. Lam, "Secure group communication using key graphs," *IEEE/ACM Trans. Netw.*, vol. 8, no. 1, pp. 16-30, Feb. 2000.
- [6] D. A. McGrew, A. T. Sherman, "Key establishment in large dynamic groups: Using one-way function trees," *IEEE Trans. Softw. Eng.*, vol. 29, no. 5, pp. 444-458, May 2003.
- [7] R. Canetti et al., "Multicast security: a taxonomy and some efficient constructions," in *Proceedings of INFOCOM '1990*, pp. 708-716, Mar 1999.

Vers un modèle des déterminants de l'usage efficace des DSSs

Application aux entrepôts de données

Nora Taibouni

Directeurs de thèse : Dr Fahima Nader ; Co-Directeur de thèse : Pr Rachid Chalal

Équipe : MSI

Résumé—Notre travail de recherche s'articule autour de l'évaluation des SIAD qui est un type particulier de système d'information¹, avec une problématique ayant comme objectif d'investiguer les facteurs déterminant l'usage réel et efficace de ce type de système. Ces facteurs contribueront dans l'amélioration du processus de développement et d'implémentation de nouveaux systèmes afin de garantir leur réussite et permettre ainsi à l'organisation de tirer profit de ces systèmes. Au titre de la deuxième année, nous avons approfondi l'analyse du domaine de l'usage des SI, qui nous a permis d'avoir une meilleure visibilité sur les possibilités de contribution. L'usage reste une problématique dans l'agenda des chercheurs SI, car son concept et son évaluation restent encore flous.

I. INTRODUCTION

L'usage est un construit qui a suscité l'intérêt des chercheurs et praticiens depuis plus de trois décennies [3]. Delone et McLean [7][8] affirment que ce concept souffre d'une conceptualisation très simpliste. L'usage des S.I. ou de la technologie de façon générale reste une problématique qui n'a pas encore trouvé consensus dans sa définition ainsi que dans sa mesure [10], et ce malgré toutes les études et recherches menées depuis les années 70s. La conceptualisation de Burton et Straub en 2006 [5] a donné un nouvel élan à la recherche sur ce concept.

L'usage est une variable essentielle dans la recherche en S.I. Elle a été utilisée en tant que variable indépendante ou intermédiaire au succès du S.I. et de la performance organisationnelle, comme elle a été utilisée en tant que variable dépendante dans les domaines d'adoption, d'acceptation et d'implémentation des technologies [5] [16]. En tant que variable dépendante, les chercheurs tentent de trouver les facteurs et les déterminants de l'utilisation des technologies, dans l'objectif d'améliorer les processus de leur conception et d'implémentation. Et en tant que variable intermédiaire, les chercheurs étudient et expérimentent les meilleures mesures pouvant capter l'utilisation qui impactent la performance individuelle et organisationnelle.

Notre travail de recherche se situe dans le contexte de l'usage en tant que variable dépendante, afin d'investiguer les antécédents de l'usage efficace des SIAD. Il y a lieu de noter que la recherche sur l'usage en particulier, et en S.I. de façon générale ne distingue pas le type du système d'information ; ceci dit, il n'y pas un domaine qui traite exclusivement de l'usage des SIAD.

Au titre de cette deuxième année, et suivant les recommandations de la commission d'évaluation de 2016, nous nous sommes focalisés sur la problématique de l'usage en particulier afin que notre étude de l'état de l'art soit la plus complète.

II. PROBLEMATIQUE

Notre problématique de recherche concerne l'évaluation des systèmes d'aide à la décision, et se focalise sur la question centrale de l'usage efficace, l'usage à valeur ajoutée, de ces systèmes, en vue d'en déterminer les facteurs contributeurs. Sans l'usage réel et efficace des S.I., les organisations ne peuvent pas tirer profit de ces systèmes et, donc leurs investissements seront inutiles [14].

Notre recherche vise un double objectif. Sur le plan pratique, elle aidera les praticiens à regarder de près les considérations de mise en place de nouveaux systèmes d'information pour un usage efficace ; et sur le plan académique elle contribuera à comprendre un peu mieux le construit de l'usage des systèmes, particulièrement les DSSs.

III. TRAVAUX REALISÉS

Durant la première année, nous avons exploré la bibliographie relative à tous les concepts liés à notre recherche : l'évaluation des DSSs, l'évaluation des entrepôts de données, l'usage et les antécédents. Nos efforts ont été dispersés sur tous ces concepts, d'autant plus que l'usage est un concept où la bibliographie est n'est pas très accessible, ancienne et très abondante. [3] affirment que entre 1992 à 2007, il y a eu un article chaque deux issues publiées dans les revues prestigieuses, *MIS Quarterly* et *Information Systems Research*.

Suivant les recommandations de la commission d'évaluation d'octobre 2016, nous nous sommes concentrés sur l'usage et comment il a été appréhendé dans la littérature, dans l'objectif de le comprendre mieux et de dégager une piste de contribution réaliste et faisable. Notre ambition était de synthétiser l'étude bibliographique sur l'usage et de la publier, travail que nous n'avons pas encore réalisé.

Nous avons résumé dans une carte conceptuelle tous les éléments d'information liés au concept (Fig. 1).

¹ Sujet de notre travail de magistère

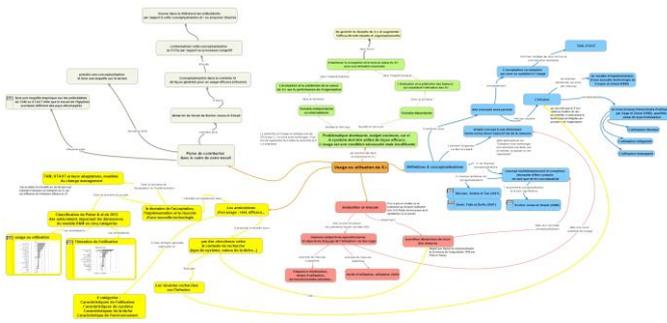


Fig.1. carte conceptuelle de System Use

IV. PERSPECTIVES

Nous nous fixons les objectifs atteignables suivants au titre de la troisième année :

1. faire une synthèse de notre étude bibliographique sur l'usage et la publier ;
2. étudier la faisabilité des pistes de contribution que nous avons dégagées avec nos encadreurs, en vue de retenir la plus significative et la développer,
3. trouver le terrain de validation de notre contribution,

V. CONCLUSION

Notre concentration, cette année, sur l'usage, nous a permis de constater que cette problématique est toujours d'actualité, malgré son ancienneté, une contribution dans ce domaine est bien faisable. Elle est inscrite et reconnue parmi les thématiques les plus pertinentes de recherche en S.I. [1], compte tenu de l'appel des chercheurs [3][10] à améliorer sa conceptualisation et sa mesure. Ajouté à cela, les échecs de mise en place de nouveaux SI rapportés par les études et la presse spécialisée.²

Cette étude nous a permis également de constater que l'usage des DSSs n'est pas appréhendé en tant que tel, il est confondu dans la recherche en S.I., c'est lors d'une étude empirique pour tester un modèle de recherche sur un DSS qu'il y a adaptation des variables de mesure. Notre travail réalisé lors du magistère sera réutilisé en grande partie.

Enfin, nous appréhendons le terrain de validation, car nous nous inscrivons dans la validation par étude empirique qui nécessite un grand nombre de répondants.

References

[1] Avison, David, and Steve Elliot. "Scoping the discipline of information systems." *Information Systems: The State of the Field* (2006): 3-18.

[2] Barr, S. H., & Sharda, R. (1997). Effectiveness of decision support systems: development or reliance effect?. *Decision Support Systems*, 21(2), 133-146.

[3] Barki, Titah, and Boffo: An Expanded Behavioral Conceptualization, *Information Systems Research* 18(2), pp. 173-192, ©2007 INFORMS

[4] Barki, Henri, Suzanne Rivard, and Jean Talbot. "A keyword classification scheme for IS research literature: an update." *Mis Quarterly* (1993): 209-226.

[5] Burton-Jones, A., & Straub Jr, D. W. (2006). Reconceptualizing system usage: An approach and empirical test. *Information systems research*, 17(3), 228-246.

[6] Chasalow, Lewis. "A model of organizational competencies for business intelligence success." Thesis at Virginia Commonwealth University, Richmond, Virginia (2009).

[7] DeLone, William H., and Ephraim R. McLean. "Information systems success: The quest for the dependent variable." *Information systems research* 3.1 (1992): 60-95.

[8] Delone, William H., and Ephraim R. McLean. "The DeLone and McLean model of information systems success: a ten-year update." *Journal of management information systems* 19.4 (2003): 9-30.

[9] Goodhue D.L., Understanding User Evaluation of Information Systems. *Management Science*, vol.41, n. 12, p. 18-27 (1995)

[10] Mclean, Ephraim R., Darshana Sedera, and Ter Chian Felix Tan. "Reconceptualizing System Use For Contemporary Information Systems." *PACIS*. 2011.

[11] Rahim, N. Z. A., Lallmahomed, M. Z., Ibrahim, R., & Rahman, A. A. (2011). A preliminary classification of usage measures in information system acceptance: a Q-sort approach. *International Journal of Technology Diffusion*, 2(4), 25-47.

[12] Roldán, J. L., & Leal, A. (2003). A validation test of an adaptation of the DeLone and McLean's model in the Spanish EIS field. *Critical reflections on information systems: a systemic approach*, 66-84.

[13] Saeed Rouhani Amir Ashrafi Ahad Zare Ravasan Samira Afshari , (2016), "The impact model of business intelligence on decision support and organizational benefits", *Journal of Enterprise Information Management*, Vol. 29 Iss 1 pp. 19 – 50

[14] Straub, D., Limayem, M., & Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management science*, 41(8), 1328-1342.

[15] Todd, P. and I. Benbasat, "Evaluating the Impact of DSS, Cognitive Effort, and Incentives on Strategy Selection", *Information Systems Research*, December 1999, pp. 356-374.

[16] Trice, A. W., & Treacy, M. E. (1988). Utilization as a dependent variable in MIS research. *ACM SIGMIS Database*, 19(3-4), 33-41.

² <http://www.silicon.fr/echec-projet-sap-coute-345-millions-euros-dhl-131307.html>

Modèles résiliables en biométrie par empreintes digitales

Hamza Djebli

Directeur de thèse : Samy Ait Aoudia

Résumé—Ce document représente un résumé des travaux réalisés durant la quatrième année en doctorat qui porte sur les modèles résiliables en biométrie par empreintes digitales. Nous allons présenter les tâches réalisées au cours de cette année, ainsi que nous donnerons nos perspectives pour l'année prochaine.

Mots clés : empreintes digitales, SIFT, biométries résiliables, transformations non inversibles

I. INTRODUCTION

Durant cette quatrième année de thèse, nous nous sommes focalisés sur la synthèse des travaux précédents et la valorisation de quelques résultats. Dans ce cadre, nous allons présenter les tâches réalisées durant cette période, et donner des perspectives pour la continuation de nos travaux.

II. PROBLEMATIQUE

La sécurité des systèmes biométriques est un aspect essentiel dans la conception de tels systèmes. Celle-ci vise à protéger la vie privée de l'utilisateur ainsi que d'empêcher toute utilisation malveillante des modèles biométriques sauvegardés. Pour cela, deux modèles générales de sécurité sont distingués : le biosystème cryptographiques, et les biométries résiliables. Notre thèse se situe dans la deuxième catégorie, où on cherche à appliquer des distorsions systématiques sur les modèles biométriques pour protéger la donnée sensible de l'utilisateur.

Dans une première partie, nous avons conçu un système d'identification par empreintes digitales, en se basant dans nos tests sur une base de données publique de référence qui s'appelle FVC (Fingerprint Verification Competition) dans ces trois versions (2000, 2002 et 2004). La méthode est basée sur la détection des emplacements des minuties, ensuite d'extraire de descripteurs SIFT pour chaque minutie, qui sont utilisés pour établir l'appariement, dans une approche inspirée du recalage d'images.

III. TRAVAUX REALISES

Dans cette année, nous avons entamé la synthèse de l'état de l'art sur l'identification par empreintes digitales et les biométries résiliables, où nous avons dégagé les chapitres suivants :

Chapitre 1 : Les systèmes biométriques

Où nous présentons les systèmes de reconnaissance de personnes de façon générale, leurs fonctionnalités et leurs applications.

Chapitre 2 : L'identification par empreintes digitales

Qui présente la structure des empreintes digitales, les méthodes d'identification et les étapes suivies par les systèmes de reconnaissance.

Chapitre 3 : Sécurité des systèmes biométriques

Qui est dédié aux types d'attaques sur les systèmes biométriques, les mesures de protection et les approches de sécurisation des systèmes biométriques, où une étude des modèles résiliables est détaillée.

Dans un autre volet, nous avons préparé un article de conférence qui représente la synthèse des travaux pratiques effectués dans la partie reconnaissance par empreintes digitales. L'article est en cours de finalisation et correction pour la soumission à la conférence SITIS 2017 (The 13th International Conference on Signal Image Technology & Internet based Systems).

Concernant la proposition d'un modèle résiliable pour la sécurisation du système de reconnaissance, les recherches sont en cours. Nous avons analysé quelques approches de hashage et de projection de descripteurs qui préservent les caractéristiques de localité entre les descripteurs de minuties (Locality Preserving Projections) qui peuvent servir comme référence.

IV. PERSPECTIVES

Comme mesures prises pour accélérer les travaux sur ma thèse, un stage de courte durée (1 mois) m'a été accordé par l'Université Abdelhamid Mehri Constantine 2, où je suis actuellement Enseignant Chercheur depuis 2 années. Le stage sera effectué dans le Laboratoire Electronique, Informatique et Image (LE2I), Université de Bourgogne, Dijon, France, et commencera en Novembre 2017. Les objectifs visés dans ce stage sont :

- Finalisation de l'état de l'art,
- Travailler sur la contribution de la thèse de doctorat dans la partie des modèles résiliables.

V. CONCLUSION

Nous avons présenté dans ce document l'état d'avancement de nos travaux sur l'identification automatique d'individus ainsi que les biométries résiliables basés sur les empreintes digitales. Malgré le fait qu'on a marqué du retard concernant les objectifs visés, nous comptons prendre des mesures sérieuses pour le rattraper.

Capitalisation des connaissances d'entreprises dans un contexte d'économie du savoir par exploitation des technologies du e-Learning

Belkacem Iskhar

Directeurs de thèse : Latifa Mahdaoui
Labo : RIIMA (USTHB), Équipe : KMEELEA

Résumé— De nos jours, l'ère de l'économie du savoir promet de régner dans un futur assez proche. D'un autre côté, le téléenseignement, ou encore le e-learning, tire ses origines des milieux professionnels en particulier celui des entreprises. Très vite d'autres aspects ont émergés et survenus comme celui de pouvoir justement capitaliser le savoir et le savoir-faire. Mieux encore, ce capital de savoir et savoir-produire peut même devenir un capital vendable. Ceci montre au fait les enjeux importants du knowledge management (KM). Le KM ou la discipline qui s'intéresse à mieux comprendre, décrire, saisir, revaloriser et progresser toutes les connaissances se trouvant à l'intérieur ou l'extérieur de l'organisation. L'objectif de ce sujet est d'asseoir et d'établir un cadre de capitalisation et de transformation des savoirs et savoir-faire de l'entreprise pour des objectifs d'économie du savoir. Cette transformation ou ce passage de connaissances imposera bien sûr l'adoption et l'usage d'une démarche qualité qui permettra de cerner ce qui est pédagogiquement valable de ce qui ne l'est pas.

I. INTRODUCTION

Un des trois piliers de l'économie du savoir, l'Intellectual Capital (IC) émerge comme une discipline scientifique à part entière sur laquelle de plus en plus se concentrent les recherches des scientifiques, se multiplient les efforts des professionnels, se focalisent les stratégies des knowledge-intensive organisations et s'orientent les politiques des gouvernements.

L'IC tourne autour de trois concept-clés : la connaissance, sa source et son apport par rapport à la valeur ajoutée de l'organisation. Ainsi, pour de multiples raisons, la source et la connaissance humaines (i.e. le Human Capital HC) se distinguent amplement des autres types de capitaux de connaissances. Le HC est à l'origine de toutes les connaissances dont une grande partie est tacite et n'appartient pas à l'organisation. Il est étudié dans plusieurs domaines, de la philosophie et l'épistémologie à la science cognitive et KM en passant par la HC Management et l'organizational learning. Chaque domaine l'aborde sous un certain angle et pour certains objectifs.

Pour les travaux de recherche pour le HC, trois remarques sont à soulever : (i) d'abord ils demeurent encore à leurs prémices étant donné la pluridisciplinarité, la complexité et la nouveauté du domaine HC ; (ii) secundo, ils tournent autour d'initiatives et d'approches de partage et de diffusion de connaissances au sein des communautés ; (iii) tertio, l'aspect pluridisciplinarité n'est pas pris en compte, i.e. que les travaux sont menés dans des domaines différents et de façon séparée, de la philosophie et l'épistémologie à la science cognitive et

Knowledge Management (KM) en passant par la HC Management et l'organizational learning.

Dans notre travail l'accent est mis sur la mise en place d'un framework conceptuel et software pour l'explicitation des connaissances tacites du HC. Afin de parvenir à notre objectif, nous avons adopté un processus méthodique et progressif de trois phases :

- Une compréhension claire aboutissant à une conception relativement pertinente à travers un state of art pour bien comprendre le domaine du HC et sa pluridisciplinarité
- Une technique ou approche permettant l'explicitation de des connaissances tacites du HC
- Un modèle de représentation supportant l'explicitation des connaissances tacites humaines

La première et la dernière phases ont été traitées dans trois travaux précédents [1, 2, 3].

Mon état d'avancement actuel se situe au niveau de la phase où l'accent sera de mettre au point des outils et conduire des tests sur un modèle expérimental ou des études de cas ou encore un formalisme pour la validation et la finalisation du framework proposé.

Il convient de préciser que l'année qui vient d'écouler, je n'ai pas progressé dans mes travaux pour diverses raisons, entre autres familiales et professionnelles. Je compte reprendre mes études et faire un pas significatif le courant de cette année.

II. FUTURS TRAVAUX

Ainsi, à la lumière de l'échéancier défini par notre encadreur, la prochaine phase consiste en la mise au point d'outils et l'accomplissement de tests sur un modèle d'expérimentation pour peaufiner notre framework d'explicitation, notamment l'approche d'explicitation et le modèle de représentation.

Nous pensons également que le recours à un formalisme se voit des plus nécessaire et pour la validation et pour la publication.

De façon plus précise, nous comptons

- Continuer les études bibliographiques et en particulier celle ayant trait à la connaissance tacite humaines et la science cognitive.

- Améliorer et développer davantage la démarche et le modèle conceptuel proposés.
- Mettre en œuvre ces deux derniers.
- Produire une publication.

- [1] B. Iskhar, L. Mahdaoui, Capitalization of enterprise's business in elearning context, International Conference on E-Learning in the Workplace, June 8th-10th 2011, New York, USA.
- [2] B. Iskhar, L. Mahdaoui, Intellectual Capital and Human Capital, State of Art and Proposal of Framework, 15th European Conference on Knowledge Management, 4-5 September 2014, Santarém, Portugal.
- [3] B. Iskhar, L. Mahdaoui, Human Capital's Explication Conceptual Framework 16th European Conference on Knowledge Management, 3-4 September 2015, Udine, Italy.

References

Vers une Holonisation des services de 'e-gouvernement' : étude des relations G to C et G to B

ATTIA Mourad

Directeurs de thèse : Latifa Mahdaoui
Labo : RIIMA (USTHB), Équipe : KMEELEA

Résumé— L'introduction des « TIC(s) » dans les divers domaines de la vie a donné lieu à l'émergence de concepts dont celui du "e-gouvernement". Les aspects collaboratifs se sont vus hissés au premier rang dans le but d'améliorer la qualité de service, avoir un meilleur rapprochement du citoyen et des partenaires socio-économiques, ... etc. Les architectures des organisations sont alors revues en faveur d'une meilleure coopération et interopérabilité afin de pouvoir fournir leurs services dans un monde de collaboration mais aussi de forte concurrence.

Dans un premier travail, nous étions intéressés à l'holonisation de procédures de services administratifs dont les outputs s'orientaient vers le citoyen. Nous avons donc pris en compte uniquement la e-administration sous la relation « G to C ».

L'objectif du travail de doctorat est de proposer un cadre d'utilisation des systèmes multi-agents holoniques pour l'assistance dans des environnements de processus administratifs complexes comme ceux engendrés par une mise en œuvre de services fournis dans un contexte de "e-gouvernance" en étudiant les possibilités offertes dans le cadre de relations « G to C » et « G to B ».

En effet, un aspect important de la e-administration est celui qui offre un ensemble de services devant couvrir les relations de type G to G, G to C et G to B. Les questions qui se posent alors sont : «Une organisation gouvernementale peut-elle architecturer ses services par une approche "Holonique" ? Quels en seraient les motivations et les gains ?".

De la réponse découleront les éventuelles applications possibles de cette technologie qui peut s'avérer utile pour améliorer entre autres la qualité de service de certaines procédures de "e-administration".

I. INTRODUCTION (HEADING 1)

Parmi les grandes problématiques et les axes de recherche dans le domaine de l'e-gouvernement est principalement l'automatisation, l'informatisation et le développement de nouveaux procédés pour soutenir le gouvernement dans la distribution des services publics et la facilité des tâches intergouvernementales.

Actuellement, les administrations publiques et leurs tâches administratives sont de plus en plus dynamiques, distribuées et complexes. Ainsi, même un processus simple (par exemple, le traitement d'une commande) peut nécessiter des transactions administratives à travers de nombreuses unités administratives (service qui est composé de plusieurs sous-services et chaque sous-service à son tour est composé de plusieurs autres sous services jusqu'au service élémentaire, c'est-à-dire non décomposable) et déclenche des interactions entre de multiples acteurs et applications logicielles.

Alors que le gouvernement électronique se réfère à la prestation de services gouvernementaux (informations, interaction et transaction) à travers l'utilisation des technologies de l'information. L'interaction entre les citoyens et les fonctionnaires administratifs se produit dans le front office, tandis que les activités d'enregistrement de traitements et d'autres ont lieu dans le back office. Cependant, plusieurs chercheurs ont constaté que la coopération back-office est un problème épuisant dans l'administration électronique en raison de problèmes d'interopérabilités.

De l'autre côté, les Systèmes Holoniques se sont développés de la nécessité de trouver la construction complète qui pourrait aider à expliquer des phénomènes sociaux. Chaque partie du système, appelée holon, est elle-même décomposable en holons. Le principe holonique veut que chaque partie soit : stable pour pouvoir faire face à des perturbations; autonome pour pouvoir agir d'elle-même afin de réaliser son objectif; coopérante pour se lier à d'autres parties afin de réaliser l'objectif commun du système. Ces systèmes font l'objet d'un courant de recherche actuellement dans le domaine de la robotique et des systèmes manufacturiers. Il est possible, en effet, de concevoir une entreprise comme une entité composée de sous entités qui, idéalement, devraient être stables et avoir des degrés d'autonomie et de coopération équilibrés .

Des questions qui peuvent être posées sont :

- Comment peut-on résoudre les problèmes d'hétérogénéité technologiques, structurels et sémantiques dans les administrations publiques?
- Comment répondre à l'évolutivité des administrations ? changement de processus métier, législation.
- Holons communiquent avec quel protocole ? Quel est le langage de requête à utiliser,....

Pour le problème d'hétérogénéité quelles sont les technologies dédiées à cette problématique ?

II. FUTURS TRAVAUX

En ce qui concerne les travaux que nous projetons pour l'année prochaine, nous comptons focaliser nos efforts et de mettre l'accent sur deux volets principaux. Le premier volet a trait à l'amélioration davantage notre framework en s'intéressant de plus aux problématiques de l'interopérabilité et de la collaboration entre les services du e-gouvernement. Le deuxième volet, quant à lui, se rapporte à la validation moyennant un formalisme.

Cela dit que nous devons solder nos travaux par une communication, par laquelle on saura déduire et repérer les points à améliorer et les points à abandonner.

Nous pensons poursuivre le plan de notre échéancier et continuer à travailler sur les BPMs . Puis nous allons adapter le modèle des RdP pour modéliser les SMAs holoniques d'une manière formelle.

Ainsi, les activités de recherche que nous comptons réaliser sont comme suit :

- Améliorer et développer davantage le processus en H que nous avons proposé, notamment sa formalisation avec un langage de formalisation ;
- Proposer une approche améliorée qui permettra à la fois l'intégration et la collaboration de modules autonomes et distribués des procédures e-gouvernementales.

- Produire une communication et/ou une publication.

Publications :

Dans ce cadre nous avons pu publier un papier qui a pour objet une proposition d'une Approche pour la l'intégration entre les services e-administratifs basés sur les systèmes multi-agents holoniques dans : « International Conference on Computer Systems and Applications Mourad, Attia ; Mahdaoui, Latifa "An approach based holonic multi-agent systems for the integration between e-administratives services"; International Conference on Computer Systems and Applications (AICCSA), ACS/IEEE 2013, Fes/Ifrane, Morocco ISSN : 2161-5322, Digital Object Identifier : 10.1109/AICCSA.2013.6616463; ©2013 IEEE

Approche d'ingénierie des connaissances en situation de collaboration intra et inter organisationnelle

Imane SID

Directeur de Thèse : Pr Abdessamad Réda GHOMARI

Équipe : MSI, LMCS

Résumé— Notre travail de recherche s'intéresse à une catégorie de processus métiers qui inclut les dimensions connaissance et collaboration et qui est par conséquent intensive en connaissances et présente de nouvelles exigences. En effet, la gestion des processus métiers a proposé des méthodes, frameworks, techniques et approches pour gérer les processus métiers. Cependant, y inclure les deux dimensions précédemment citées rend ces processus plus complexes et pose de nouveaux défis scientifiques objet de cette thèse de doctorat.

I. INTRODUCTION

La gestion des processus métiers [1] et la gestion des connaissances [2] [3] sont deux domaines de recherche distincts à partir desquels a émergé une catégorie de processus métier spécifique rassemblant les trois dimensions : processus métier, connaissance et collaboration.

L'importance de ces processus est appuyée par une analyse du grand bureau de conseil Gartner: "By 2014, 40% of business managers and knowledge workers in global 2000 enterprises will use comprehensive business process models to support their daily work up from six percent in 2009" .

II. PROBLEMATIQUE

La gestion des processus métiers a apporté jusqu'à ce jour des approches, des méthodes et des techniques permettant la modélisation, l'exécution et le diagnostic des processus métiers impliquant des acteurs humains, organisations, ressources, etc.

Comme résultat, beaucoup de langages permettant la modélisation de ces processus sont apparus :

- *Des langages formels* tels que les réseaux de pétri sui sont considérés comme des modèles mathématiques ;
- *Des langages conceptuels* dont le BPMN (Business Process Management Notation) est le plus utilisé ;
- *Des langages d'exécution* dont le BPEL (Business Process Execution Language) est le plus connu.

Beaucoup de systèmes de gestion des processus métiers BPMS (Business Process Management System) ont également vu le jour afin de permettre et de supporter la modélisation, l'exécution et le diagnostic de ces processus.

Parmi les systèmes de gestion des processus métiers les plus connus et utilisés dans le domaine académique et industriel, nous citons [4] :

- *BizAgi Xpress*: Offrant un environnement graphique pour le PBM permettant de modéliser, d'exécuter et d'améliorer les processus métiers ;

- *Bonita* : Une solution complète de BPM incluant trois modules (Bonita Studio, Bonita User Experience and Bonita Execution Engine). Elle implémente les éléments les plus importants définis dans le BPMN 2.0 ;

- *Oracle BPM Suit 11g* : Une solution fournissant des outils permettant aux utilisateurs de modéliser et de gérer leurs processus métiers. Elle regroupe deux outils : JDeveloper-based BPM Studio et web-based Process Composer.

- *ADONIS* : est un ensemble d'outils pour la modélisation des processus. Sa philosophie est l'amélioration continue des processus. Des interfaces telles que XPD, BPMN et XML constituent des implémentations réussies du modèle ADONIS.

Cependant les processus métiers incluant les dimensions connaissances et collaboration présentent des caractéristiques qui font que les méthodes, techniques, approches et frameworks existants ne les gèrent plus de manière efficace.

III. TRAVAUX REALISES

A. Etat de l'art sur trois domaines d'intérêt (1ère année de thèse)

Un état de l'art sur chacun des trois domaines de recherche est nécessaire: la gestion des processus métiers, la gestion des connaissances et la collaboration afin de nous familiariser avec les concepts de bases et identifier les rapprochements entre chacun des trois domaines.

Un processus métier est un ensemble d'activités ordonnancées, ayant un début et une fin [5]. Il permet de décrire la manière dont est organisé le travail au sein d'une organisation et crée de la valeur ajoutée pour le client/marché [2].

La discipline qui s'intéresse aux méthodes, techniques et logiciels permettant de concevoir, exécuter, contrôler et analyser ces processus opérationnels impliquant des acteurs humains, organisations, applications, documents et autres ressources d'information est dite *gestion des processus métiers* (BPM) [1].

La connaissance est de l'information combinée à de l'expérience, du contexte, de l'interprétation et de la réflexion [3]. Elle est du savoir, savoir-faire, de l'expérience, voire du savoir être » [6].

La collaboration est un processus impliquant différents agents qui peuvent avoir différentes perceptions d'un problème et qui sont engagés dans un processus à travers lequel ils cherchent à trouver une solution en dépassant leurs expertises et visions individuelles [7].

Nous avons par conséquent effectué un état de l'art sur la gestion des processus métier, la gestion des connaissances et le domaine de la collaboration.

B. Etude de la collaboration dans les processus métiers (2^{ème} année de thèse)

Nous nous sommes intéressés à la collaboration dans les processus métiers à travers une 1ère communication intitulée: «*Cartographie des situations de collaboration dans les processus métiers: Application à la gestion des stages conventionnés en entreprise* » [8] dans laquelle nous avons identifié des situations de collaboration génériques dans la gestion des processus métiers entre les différents acteurs qui peuvent y être impliqués comme illustré dans *table 1* :

TABLE I. SITUATIONS DE COLLABORATION POSSIBLES /PHASES DU BPM

Acteurs / Phases du BP	Exécuteurs	Modélisateurs	Experts métiers
Conception			
Configuration		Validation du modèle du BP réalisé	
Exécution	Optimisation des tâches		
Diagnostic	Amélioration et correction des failles du modèle actuel		

C. Etude de la dimension connaissance dans les processus métiers (2^{ème} année de thèse)

Nous nous sommes par la suite intéressés à la gestion des connaissances dans les processus métiers. Une seconde communication a été réalisée, intitulée : « *Knowledge cartography and capitalization in collaborative business processes* » [9] dans laquelle nous avons identifié et cartographié les connaissances générées et utilisées par chacune des phases du cycle de vie du processus métier comme illustré dans *table 2* :

TABLE II. CONNAISSANCES UTILISEES ET GENEREES PAR LES PHASES DU CYCLE DE VIE DU PROCESSUS METIER.

Phases du BP	Connaissances utilisées	Connaissances générées
Modélisation	Connaissances résultant des activités du CPI/BPR	Connaissances sur le modèle actuel
Configuration	Connaissances sur l'environnement organisationnel	Connaissances sur la configuration actuelle
Exécution	Connaissances comme ressource nécessaire à l'exécution des tâches	-Nouvelles connaissances générées par les acteurs après exécution des tâches. -Connaissances sur l'instance actuelle
Diagnostic	Connaissances sur l'exécution des instances : temps d'exécution, coût,...	Indicateurs de performance/qualité du BP

D. Proposer un support efficace aux exigences que présentent les processus métiers (3^{ème}, 4^{ème} et 5^{ème} années de thèse)

Nous nous sommes concentrés sur l'analyse des caractéristiques qui émergent des dimensions connaissance

et collaboration dans les processus métiers afin de proposer un support efficace à ces derniers.

Nous avons effectué dans ce contexte un séjour scientifique de 11 mois (Octobre 2015 - Août 2016) à l'Institut de bases de données et systèmes d'information de l'Université d'ULM en Allemagne dans le cadre de la bourse PNE. Ce stage a été l'occasion d'échanger avec les chercheurs affiliés à l'institut et d'identifier des pistes d'idées de publications.

Nous avons rédigé et soumis un article de journal et nous sommes entrain de structurer et rédiger les chapitres du mémoire de la thèse.

IV. PERSPECTIVES

Dans les prochains mois nous continuerons la rédaction du mémoire de la thèse.

V. CONCLUSION

Nous avons effectué un état de l'art sur les domaines de la gestion des processus métiers, la gestion des connaissances et la collaboration.

Nous avons analysé les processus métiers et avons identifié des caractéristiques mal supportées par les approches et outils existants. Nous avons rédigé et soumis un article de journal dans ce contexte.

Nous continuerons la rédaction du mémoire de la thèse dans les prochains mois.

References

[1] W. P. Aalst, A. M. Hofstede, and M. Weske, "Business Process Management: A Survey," in Business Process Management. vol. 2678, W. P. Aalst and M. Weske, Eds., ed: Springer Berlin Heidelberg, 2003, pp. 1-12.

[2] M. Hammer and J. Champy, Reengineering the corporation: Harper Business, 1988.

[3] T. H. Davenport and L. Prusak, Working knowledge: How organizations manage what they know: Harvard Business Press, 2000.

[4] L. Hui, Y. Lemrabet, D. Clin, and J. Bourey, " In Research Challenges in information Science (RCIS), F," presented at the Research Challenges in information Science (RCIS), Fifth International Conference on, 2011.

[5] T. H. Davenport, Process innovation: reengineering work through information technology: Harvard Business Press, 1993.

[6] I. Nonaka, H. Takeuchi, and H. Takeuchi, The knowledge-creating company: how Japanese companies create the dynamics of innovation. , 1995.

[7] C. Shah, "Collaboration," in Collaborative Information Seeking. vol. 34, ed: Springer Berlin Heidelberg, 2012, pp. 11-24.

[8] I. Sid, A. R. Ghomari, and A. Benhouhou, "Cartographie des situations de collaboration dans les processus métiers: Application à la gestion des stages conventionnés en entreprise," 2ème Conférence Francophone sur les Systèmes Collaboratifs SYSCO'2014, Hammamet, Tunisie, 2014.

[9] I. Sid and A. R. Ghomari, "Knowledge cartography and capitalization in collaborative business processes ", 4eme Colloque International ISKO-Maghreb'2014: Concepts and tools for knowledge management, CERIST, Alger, Algérie, 2014.

Détection de communautés dynamiques par une approche d'analyse des réseaux sociaux

Boudebza Souâad

Directeurs de thèse : Nouali Omar, Azouaou Faïçal

Équipe : EIAH

Abstract—La détection de communautés est une problématique majeure dans le domaine d'analyse des réseaux sociaux (RS). La majorité des travaux dans ce domaine portent sur des communautés statiques. Or, les RS sont dynamiques. L'objectif de cette thèse est de proposer une approche pour la détection de communautés dynamiques.

I. INTRODUCTION

A l'instar d'Internet, les réseaux sociaux ont émergé de manière surprenante. En effet, ils sont parvenus à faire exploser le nombre d'utilisateurs et la quantité d'information échangée. Par exemple, sur le réseau social Twitter, on a 900 millions présents, publiant environ 400 millions de Tweets par jour.

Avec cette prolifération des réseaux sociaux et leur évolution continue, leur analyse suscite de vifs intérêts [1]. La détection de communautés est une problématique majeure dans le domaine de l'analyse des réseaux sociaux. Elle est fondée principalement sur la théorie de graphes[2]. Elle consiste à former des groupes de sorte que les nœuds au sein d'un même groupe soient connectés d'une manière dense. Par conséquent, chaque groupe ainsi identifié possède une forte connectivité à l'intérieur et une faible connectivité avec les autres groupes.

La littérature scientifique concernant la détection de communautés est très abondante. La majorité des travaux dans ce domaine portent sur des communautés statiques. Or, les réseaux sociaux sont en fait dynamiques, ils évoluent au cours du temps. Par exemple, à tout moment, des utilisateurs peuvent rejoindre ou quitter le réseau, une relation peut être établie entre deux personnes, un membre peut ajouter ou supprimer du contenu (document, image, etc.) sur le réseau, etc. Les approches de détection de communautés dynamiques s'orientent dans deux grandes directions. La première direction considère un réseau dynamique comme une succession de réseaux statiques et elle prend davantage de réutiliser les algorithmes de détection statiques sur chaque instantanée du réseau. Celle-ci présente toutefois une faible efficacité et beaucoup d'informations sur l'historique de communautés sont négligées. La seconde direction utilise directement les graphes temporels [3], ici l'évolution du réseau est considérée comme une suite de modifications sur le réseau (ajout et/ou suppression des nœuds et des liens). Cette approche est capable de traiter un réseau au fur et à mesure de son évolution.

Objectif de cette thèse est de proposer une nouvelle approche basée réseaux temporels pour la détection de communautés dynamiques les dans les réseaux sociaux.

II. PROBLÉMATIQUE

L'objectif de cette thèse est de proposer une nouvelle approche pour la détection de communautés dynamiques dans

des réseaux sociaux. La dynamique de ces réseaux peut être traitée de deux manières:

- 1) soit comme une série d'instantanés qui sont tous des graphes statiques représentant l'état du réseau à un instant donné
- 2) soit comme un réseau temporel constitué de nœuds et de liens auxquels on associe des séquences d'intervalles, correspondant aux intervalles de temps durant lesquels le nœud ou le lien est présent dans le réseau.

Les approches adoptant le premier modèle de présentation ont l'intérêt de pouvoir réutiliser les méthodes statiques, sur chacun des instantanés on applique un algorithme classique, puis on fait correspondre les communautés trouvées dans un instantané avec celles de l'instantané précédent (suivi de communautés). Les problèmes qui soulèvent dans ces approches sont : la complexité qui augmente proportionnellement avec le nombre de pas de temps et la taille du réseau et le problème d'instabilité, en effet, les résultats de détection pourraient être très différents pour des instantanés similaires.

Dans les approches travaillant sur des réseaux temporels les communautés existantes sont modifiées à chaque modification sur le réseau. Il n'y a plus ici de problème d'instabilité.

Outre la question de représentation de l'évolution de réseau, d'autres difficultés pourraient se présenter à la détection de communautés dynamiques :

- 1) le recouvrement de communautés, c-à-d la possibilité d'appartenance d'un nœud à plusieurs communautés. Cette caractéristique est très courante dans les réseaux réels, comme les réseaux sociaux.
- 2) le cycle de vie de communautés: L'évolution des communautés se résume dans les opérations suivantes: naissance, mort, croissance, contraction, fusion et division. Ces opérations ne sont pas toujours prises en considération.

L'approche que nous allons proposer doit donc répondre à ces différents enjeux.

III. TRAVAUX RÉALISÉS

La majeure partie des deux premières années de thèse a été consacrée à la recherche bibliographique. Nous avons réalisé un état de l'art sur les approches de détection de communautés dans les grands graphes, et en particulier dans les graphes dynamiques. Nous avons pu constater qu'il existe deux grandes directions pour la détection de communautés dynamiques. Les

approches de la première direction travaillant sur des instantanés du réseau et souffre toutefois du problème d'instabilité. Les approches de la deuxième direction utilisent les réseaux temporels [3] et elle est capable de traiter un réseau au fur et à mesure de son évolution. Nous nous sommes orientés vers cette seconde direction. Dans un premier temps nous avons proposé une amélioration de l'algorithme iLCD proposé par Cazabet [4]. Il s'agit d'un algorithme récent et performant, auquel nous avons intégré l'aspect hiérarchique, pour permettre une détection à plusieurs niveaux (c.à.d. détecter les communautés de communautés). En outre, une revue de littérature a été menée sur les approches d'analyse des réseaux sociaux. Nous avons examiné les modèles pour la représentation des RS afin d'intégrer la dimension temporelle et de supporter la détection de communautés dynamiques. Nous avons par la suite proposé un modèle sémantique et temporel pour la représentation des réseaux sociaux. Cette proposition a été publiée dans la conférence FiCloud 2015 [5].

Durant la troisième année de thèse, nous avons proposé un nouvel algorithme de détection de communauté, basé sur la notion de cliques et de graphes temporels. Nous avons opté pour une approche à base de cliques pour permettre une détection de communautés recouvrantes. En outre, l'utilisation des graphes temporels nous a permis de suivre le réseau au fur et à mesure de son évolution. Cet algorithme a été implémenté et testé. Les premiers résultats des expérimentations sur des réseaux de test statiques et dynamiques à petite échelle étaient encourageants, mais nous avons estimé qu'il est nécessaire d'améliorer l'outil pour résoudre les deux problèmes rencontrés, celui du passage à l'échelle et celui de la qualité de découpage.

Durant la quatrième année, nous nous sommes focalisé sur l'amélioration et la validation de notre proposition. Nous avons réalisé plusieurs expérimentations à la fois sur des réseaux synthétiques et des réseaux sociaux réels à large échelle. Les résultats obtenus montrent l'efficacité de notre méthode. Un article est soumis sur notre proposition [6].

IV. PERSPECTIVES

Les objectifs soulignés durant la cinquième année de thèse sont les suivants :

- 1) Rédiger la thèse.
- 2) Effectuer les éventuelles corrections et modifications sur l'article

V. SÉJOURS SCIENTIFIQUES

Nous avons bénéficié de deux séjours scientifiques :

- 1) Durant l'année 2014/2015: un séjour d'un mois à l'Institut de Recherche en Informatique de Toulouse (IRIT).
- 2) Durant l'année 2015/2016: un séjour d'un mois au Laboratoire d'Informatique de Paris 6 (LIP6) avec l'équipe Complex Networks.

VI. ACTIVITÉ CONNEXES

Parallèlement à mes activités de recherche, je suis maître assistante au département d'informatique à l'université Mohammed Seddik Ben Yahia de Jijel. J'ai encadré trois travaux de master qui sont articulés autour de la problématique de détection de communautés.

REFERENCES

- [1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge university press, 1994.
- [2] F. Harary and R. Z. Norman, "Graph theory as a mathematical model in social science," Ann Arbor, Mich. : Xerox University Microfilms, Tech. Rep., 1976.
- [3] P. Holme and J. Saramäki, "Temporal networks," *Physics reports*, vol. 519, no. 3, pp. 97–125, 2012.
- [4] R. Cazabet, "Détection de communautés dynamiques dans des réseaux temporels. (detection of dynamic communities in temporal networks)," Ph.D. dissertation, Paul Sabatier University, Toulouse, France, 2013. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-00874017>
- [5] S. Boudebza, F. Azouaou, and O. Nouali, "Ontology-based approach for temporal semantic modelling of social networks," in *3rd International Conference on Future Internet of Things and Cloud, FiCloud 2015, Rome, Italy, August 24-26, 2015*.
- [6] S. Boudebza, R. Cazabet, F. Azouaou, and O. Nouali, "Olcpm: An online framework for detecting overlapping communities in dynamic social networks," *Computer Communications*, 2017.

La théorie des jeux pour la résolution du problème d'affectation de fréquences dynamique dans un réseau GSM

Laidoui FATMA

Directeurs de thèse : Professeur Mme Benbouzid-Si Tayeb Fatima et le docteur Mme Bessedik Malika

Équipe : Optimisation Combinatoire

Année d'inscription : 2015-2016

Abstract—La résolution du problème d'affectation de fréquences dans les réseaux GSM prend toujours l'attention des chercheurs, ce qui conduit l'existence de plusieurs modèles dans la littérature pour faire face à ce problème. Notre objectif consiste plus particulièrement à essayer de résoudre le problème d'affectation de fréquences dynamique en utilisant l'outil mathématique la théorie des jeux.

I. INTRODUCTION

Les réseaux mobiles basant sur le standard GSM sont les technologies les plus répondues dans ces dernières décennies. Ce type de réseau utilise les ondes radio pour la communication. Chaque opérateur mobile dispose un nombre limité de fréquences. En conséquence et afin d'essayer d'augmenter la capacité du réseau, la réutilisation de ces fréquences est inévitable. Le problème d'affectation de fréquences (FAP) est classé dans la catégorie des problèmes d'optimisation combinatoire [4], [3], [1]. Afin d'essayer de résoudre ce problème, plusieurs approches ont été proposées dans la littérature. parler sur la théorie des jeux

Le document est organisé comme suit: dans la section II, nous exposons notre problématique. Dans la section III, nous présentons les travaux réalisés. Dans la section IV, nous proposons des perspectives à réaliser. Et finalement dans la section V, nous terminons par une conclusion.

II. PROBLÉMATIQUE

Notre travail consiste à proposer une solution pour résoudre le problème d'affectation de fréquences dynamique dans les réseaux GSM en utilisant la théorie des jeux.

III. TRAVAUX RÉALISÉS

A. Proposition d'une solution

Avant de déterminer le type de jeux utilisé pour résoudre notre problématique, il faut tout d'abord savoir quels sont les éléments de base du jeu. En effet, un jeu comprend un ensemble de *joueurs* ayant chacun un ensemble d'*actions* possibles et une fonction de *gain* qui est une valeur associée pour chaque combinaison possible d'actions.

1) *Les joueurs*: Les joueurs dans notre modèle sont les fréquences. Le choix de ces joueurs est justifié par : (i) les acteurs principaux dans le jeu sont les fréquences. En effet, tout le jeu dépend de l'allocation des ces fréquences dans leurs canaux. (ii) L'information sur le nombre des fréquences dans les réseaux mobiles est connue.

2) *Les actions*: Une action est le fait d'affecter une fréquence à un canal. Un joueur peut choisir plusieurs canaux si les contraintes des réseaux GSM sont respectées.

3) *Les règles du jeu*: Les règles du jeu sont les contraintes de séparation. Cette information est connue à l'avance.

4) *Le gain*: Notre objectif principal est de résoudre le problème d'affectation de fréquences tout en respectant la qualité de service des réseaux mobiles. En conséquence, la réflexion sur le choix de la fonction gain se porte sur les deux alternatives suivantes: (i) minimiser les interférences, donc minimiser le bruit dans la communication ; (ii) ou minimiser le temps de bout en bout, donc l'utilisateur ne va pas trop attendre pour que sa demande d'appel atteigne sa destination. Nous procédons par élimination, la première alternative est à éliminer, car nous pensons que cet objectif est inclus dans les contraintes de séparation. Ainsi, la fonction d'utilité de notre jeu est de minimiser le temps de bout en bout.

5) *La solution d'un jeu*: Une solution du jeu consiste à trouver un état stable de système (équilibre de système).

6) *Déterminer le type de jeu*: Après avoir fixé les éléments de base du jeu dans la première phase, la deuxième phase consiste à trouver le type du jeu qui pourra résoudre notre problématique. Dans notre jeu, le gain d'un joueur n'implique pas systématiquement la perte de l'autre joueur, donc le jeu est un jeu à **somme constante**. Le but de jeu est de trouver l'affectation optimale en un minimum de temps, donc le jeu ne doit pas se dérouler à l'infini. En conséquence, le jeu est un jeu **discret** non continue. Dans notre jeu, chaque joueur (fréquence) n'a pas besoin de savoir les stratégies des autres joueurs. Ainsi, le jeu est un jeu **non-coopératif**. Les joueurs jouent d'une manière **séquentielle** car s'ils jouent simultanément, on risque de choisir deux fréquences dans des zones causant une interférence. La stratégie de chaque joueur n'est pas forcément la même pour tous les joueurs donc le jeu est **asymétrique**. Dans notre jeu, il existe plus de deux joueurs, donc le jeu est un jeu à **n-joueurs**.

Notre jeu est donc un jeu à Somme Constante, Discret, Non-coopératif, Asymétrique avec Multiples joueurs, nous l'appelons **CSDNcAM** (Constant Sum Non-cooperative with Asymmetric Multiple players game)

B. Critiquer l'approche proposée

Au début, cette approche paraît idéale pour résoudre notre problématique. Cependant, si nous analysons bien les éléments

de base du jeu, nous remarquons que choisir les fréquences comme joueurs pourraient rendre la solution non flexible à l'ajout d'autres joueurs, c'est à dire l'ajout d'une nouvelle fréquence. Certainement, cette solution pourrait résoudre le problème d'affectation de fréquences dynamique en prenant en considération le principe de réutilisation des fréquences et le problème d'interférences (en respectant les contraintes de séparation). En revanche, il paraît que l'ajout d'une nouvelle fréquence nécessite toute une phase d'adaptation. D'où notre première réflexion qui se porte sur le changement des éléments de base du jeu.

Ajoutons à ceci, la théorie des jeux évolutionnaire (EGT) est une autre perspective que la théorie de jeu classique utilisée pour les réseaux sans fil. La différence entre GT classique et EGT est que chaque joueur dans GT classique choisit d'une manière rationnelle une stratégie parmi un ensemble *statique* de stratégies. En conséquence, il est fondamental dans GT classique que chaque joueur analyse la stratégies des ses adversaires avant de choisir sa propre stratégie. En contrepartie, dans GT évolutionnaire (EGT) qui est une forme plus récente de GT, les joueurs ne sont pas tous rationnels. En effet, les joueurs ont des informations limitées sur les choix disponibles et leurs conséquences. Aussi, les stratégies ne sont pas statiques dans EGT (les stratégies *évoluent*). Les joueurs ont une stratégie privilégiée qui est continuellement classée avec d'autres stratégies alternatives [5]. Ainsi, nous pensons qu'il faut aussi étudier cette alternative.

Aussi, nous distinguons qu'il existe un type de jeu coopératif spécifique appelé jeu *Stackelberg* (SG) [9]. L'idée initiale derrière un SG est qu'un joueur, appelé *leader*, a le droit de faire la première action. Cette action est sélectionnée pour que le leader optimise sa récompense après avoir observé les stratégies des autres joueurs (*followers*). Ensuite, le leader annonce sa stratégie privilégiée pour les followers. Ces derniers observent l'action du leader et adaptent leurs stratégies afin de minimiser leur propre coût. Après cette étape, les followers annoncent leurs stratégies à nouveau au leader. Ainsi, nous pouvons définir un SG comme un modèle séquentiel qui analyse l'interaction entre un leader et un ensemble de followers afin qu'au moins un objectif modèle spécifique puisse être atteint. Le but principal de ce type de jeu est de découvrir l'équilibre Stackelberg(SE) donné par le vecteur (Strategy_leader; Strategy_follower) [5]. Nous trouvons dans [2] que le jeu Stackelberg est utilisé dans différents types de réseau y compris les réseaux cellulaires.

C. Approche proposée

Dans notre nouvelle approche nous basons sur le jeu *Stackelberg* pour modéliser notre jeu. Nous avons besoin principalement de deux joueurs J_1 qui est le *leader* responsable de l'affectation des fréquences aux émetteurs et un autre joueur J_2 qui est le *follower* responsable à l'ordonnement des demandes de fréquences en attente. Lorsque il y a une demande de fréquence, le leader vérifie si cette fréquence est disponible, si c'est le cas, il répond à la demande. Dans le cas inverse, le leader décide d'envoyer cette information de demande de fréquence au joueur follower J_2 et lui demande de l'envoyer la demande de fréquence en attente. Ce dernier, selon une stratégie de classement des demandes enregistre cette information. Le joueur follower J_2 applique le principe de FIFO pour ordonnancer les demandes de fréquences en

attente et envoie au leader J_1 la demande de fréquence en attente. L'objectif du J_1 est d'optimiser le temps de réponse à la demande de fréquence. L'objectif du J_2 est de réduire le nombre de fréquences en attente. La stratégie du J_1 est l'affectation de la fréquence si c'est disponible. Solliciter J_2 sinon. La stratégie du J_2 est le classement des fréquences en attente selon leur ordre d'arrivée. La figure 1 récapitule les quatre étapes de notre jeu Stackelberg.

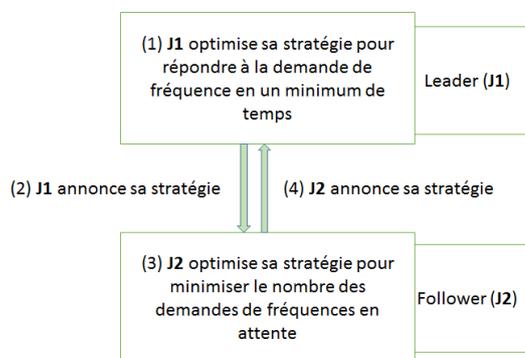


Fig. 1: Les quatre étapes dans le jeu Stackelberg proposé

IV. PERSPECTIVES

Selon l'étude effectuée dans la section III-B, nous déterminons les perspectives suivantes: (1) Implémenter et tester la solution initiale. (2) Garder le modèle du jeu initial et changer les ingrédients de base du jeu. (3) implémenter et tester le jeu proposé dans (2). (4) Proposer un autre modèle basé sur l'EGT. (5) Implémenter et tester le modèle proposé dans (4). (6) Implémenter et tester la solution proposée dans la section III-C. (7) Envisager plusieurs stratégies pour les joueurs. (8) Continuer la rédaction.

V. CONCLUSION

Dans ce présent résumé, nous avons rappelé notre problématique. Ensuite, nous avons proposé une solution en détaillant chaque élément de base constituant le jeu. Après, nous avons critiqué le modèle proposé, ce qui nous a conduit à une nouvelle approche de résolution. Nous avons enchaîné par détailler le nouveau modèle du jeu. Ceci, nous a permis de déterminer nos perspectives. Finalement, nous avons terminé par une conclusion.

REFERENCES

- [1] A.M.C.A Koster. Frequency assignment problem, models and algorithms.1999.
- [2] Hoang, Dinh Thai and Lu, Xiao and Niyato, Dusit and Wang, Ping and Kim, Dong In and Han, Zhu. Applications of repeated games in wireless networks: A survey. IEEE Communications Surveys & Tutorials. 2015.
- [3] Mathar R., Mattfeldt J.. Channel assignment in cellular radio networks. IEEE Transactions on Vehicular Technology,1993.
- [4] Kunz D.. Channel assignment for cellular radio using neural network. IEEE Transactions on Vehicular Technology, 1991.
- [5] Moura, Jose and Hutchison, David. Survey of Game Theory and Future Trends with Application to Emerging Wireless Data Communication Networks. arXiv preprint arXiv:1704.00323. 2017.

- [6] Octave Boussaton. Application de la théorie des jeux à l'optimisation du routage réseau - solutions algorithmiques. Thèse de doctorat. Université Henri Poincaré, France. 2010.
- [7] R Duncan Luce, Howard Raiffa. Games and decisions : introduction and critical survey. New York : Dover Publications,1957.
- [8] Sudip Misra, Barun Kumar Saha, Sujata Pal. Opportunistic Mobile Networks Advances and Applications. Springer International Publishing Switzerland. 2016.
- [9] Xiumin Wang and Xiaoming Chen and Weiwei Wu and Ning An and Lusheng Wang. Cooperative Application Execution in Mobile Cloud Computing: A Stackelberg Game Approach. IEEE Communications Letters. 2016.

Systèmes multi-agents ultra large échelle : contributions à l'introduction dynamique de systèmes dans un système de systèmes

Abdelhakim Hamzi

Directeurs de thèse : Professeur M. Koudil

Équipe : CoDesign

Résumé—Notre travail repose sur l'application des approches multi-agents pour la modélisation et l'intégration dynamique d'un sous-système dans un système de systèmes. Pour cela, nous proposons une représentation formelle basée sur les SMA. Pour résoudre le problème de sélection de la meilleure combinaison des systèmes à intégrer nous avons utilisé les algorithmes génétiques multi-objectifs.

I. INTRODUCTION

Un système de systèmes (SdS) est généralement un système de grande taille composé de sous-systèmes, appelés composants systèmes (CS), imbriqués les uns dans les autres, autonomes, souvent hétérogènes, pouvant communiquer et s'adapter de manière locale ou globale pour répondre aux changements de leur environnement. L'ingénierie de tels systèmes fait l'objet d'importants efforts en termes de travaux de recherche. Le problème d'intégration de SdS est un problème majeur à cause de la complexité et l'hétérogénéité des systèmes à intégrer, les comportements émergents, et la difficulté de maintenir la collaboration et l'interopérabilité entre les CS.

Les systèmes multi-agents (SMA) sont une approche privilégiée pour s'intéresser aux systèmes complexes grâce à leur démarche totalement décentralisée. Notre objectif est d'appliquer les techniques de ce domaine pour répondre au problème de l'intégration dynamique d'un sous-système dans un SdS. Pour cela un modèle multi-agents intégrable sera présenté ainsi qu'une solution pour l'intégration de systèmes dans un SdS. Afin valider notre travail nous proposons d'intégrer plusieurs réseaux de capteur dans une seule plateforme ou système.

II. PROBLÉMATIQUE

Le concept de SdS reste aujourd'hui trop peu formalisé, entre autres il n'existe pas encore de modèle généraliste permettant de le contrôler en assurant le respect de ses caractéristiques fondamentales. L'intérêt principal des SdS repose sur un double paradoxe. D'une part, le modélisateur veut s'assurer le contrôle du SdS tout en laissant une grande liberté de décision aux CS. D'autre part, la conception d'un SdS se fait par un mélange d'intégrations horizontales et verticales. L'intégration horizontale revient à ajouter de nouveaux systèmes qui vont interagir avec ceux déjà présents dans le SdS. L'intégration verticale revient à faire apparaître plusieurs niveaux hiérarchiques de modélisation.

Pour répondre aux besoins de modélisation des SdS, l'utilisation de l'approche multi-agents est une solution appropriée. L'organisation, les composants et les interactions entre les composants dans tels systèmes complexes nous mènent à utiliser l'approche AEIO (Agent-Environnement-Interaction-Organization) en se basant sur la caractéristique récursive du modèle. Donc le problème d'intégration d'un système dans un SdS sera reformulé par ces deux questions: Comment gérer l'arrivée d'un nouveau système dans un SMA récursif en cours de fonctionnement ? et comment l'intégrer

dynamiquement si le besoin s'en fait sentir ? Pour cela, il est nécessaire de proposer une description sémantique des CS à intégrer (méta-modèles, propriétés) et définir les décisions que le système global doit prendre à partir de la description sémantique des CS (Choix du niveau d'introduction et impact sur le fonctionnement multi-niveau du système).

III. ETAT DE L'ART

1. Généralités sur les SdS: Nous avons introduit le concept des SdS et leurs caractéristiques. Une étude sur les applications et les techniques d'ingénierie des SdS est faite.
2. La modélisation des SdS: Nous avons présenté les concepts de base de l'ingénierie dirigée par la modélisation et les différents langages et standards de modélisation. Ensuite nous avons parcouru les différents approches et techniques de modélisation. Enfin, nous avons cité les travaux dans la littérature en ce qui concerne la modélisation des SdS.
3. Approches et techniques pour l'intégration des SdS: Nous avons clarifié le concept d'intégration de SdS. Ensuite, nous avons montré les défauts de ce problème et les différentes approches et technologies utilisées. Enfin un parcours des travaux sur ce domaine est fait.
4. Les SMA: Les notions de base des agents et systèmes multi-agents sont introduits. Ensuite les grandes approches des SMA et classification des architectures de ces dernières sont présentées. Enfin, une étude comparative des méthodologies des SMA est faite.

IV. CONTRIBUTION

1. Méta-modèles pour les systèmes de systèmes

Nous avons proposé un méta-modèle pour la modélisation des SdS. Le méta-modèle est basé sur quatre axes:

- A. *Méta-modèle « Agent-Organization »*: Il exprime les différents éléments et leurs rôles dans le système et comment ils forment des groupes pour atteindre un objectif.
- B. *Méta-modèle de l'environnement* : Il définit les propriétés du monde dans lequel une entité (agent ou objet) existe et peut fonctionner.
- C. *Le méta-modèle « Task-Goal »*: Il décrit les relations entre les buts et les tâches nécessaires pour les atteindre ainsi que leurs effets sur le système et l'environnement.
- D. *Le méta-modèle « Interaction »*: pour exprimer les interactions et les communications entre les agents et les objets de l'environnement.

2. Mécanisme d'intégration de SdS

Notre problème consiste à intégrer plusieurs systèmes qui sont déjà opérationnels pour accomplir un but commun.

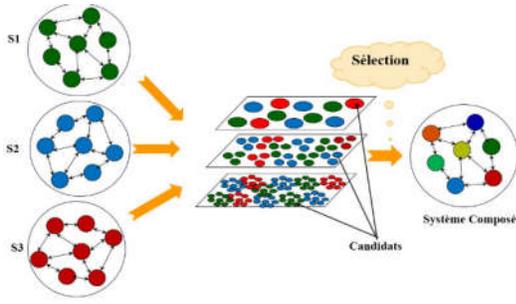


Fig. 1. Problème de d'intégration de systèmes

L'opération d'intégration se fait en quatre phases :

1. La définition des préférences du nouveau système à composer en utilisant la description sémantique des CS.
2. La sélection: permet de choisir les meilleurs éléments appartenant aux systèmes à intégrer qui vont composer le nouveau système en répondant aux besoins et en assurant un certain niveau de QoS.
3. Le déploiement: reconfiguré les éléments choisis pour jouer un rôle dans le nouveau système composé.
4. L'exécution: exécuter le nouveau système et observer les paramètres de performance.

Chaque agent joue un rôle particulier dans le système pour atteindre un but global. Pour jouer ce rôle, l'agent doit avoir des capacités. La capacité est réalisée en exécutant des tâches selon un plan bien défini. Une tâche peut être réalisée par plusieurs méthodes en utilisant des ressources. La méthode réalise une tâche par l'exécution de plusieurs opérations élémentaires.

La description des agents.

Les propriétés de l'agent PA

PA = (Type , State, Ressources , Capabilities, Roles)

Type : les types de l'agent

$$Type = \bigcup_{i=1}^n tp_i \quad \dots \dots \dots (1)$$

State : l'état de l'agent qui peut être mesuré en fonction de certains paramètres.

$$State = \bigcup_{i=1}^n \left(pr_i, \bigcup_{j=1}^m val_j \right) \quad m = 1, 2, \dots, k \quad \dots \dots (2)$$

Ressources: les ressources disponibles pour l'agent.

Ressources = ((rs₁,value), (rs₂,value), ..., (rs_n,value))

$$Res = \bigcup_{i=1}^n \left(res_i, \bigcup_{j=1}^m val_j \right) \quad m = 1, 2, \dots, k \quad \dots \dots (3)$$

Capabilities : Les capacités de l'agent en termes de tâches qu'il peut accomplir avec un degré de qualité de service si certaines conditions sont vérifiées.

$$Capabilities = \bigcup_{i=1}^n cp_i \quad \text{Où} \quad \dots \dots (4)$$

$$cp_i = \bigcup_{j=1}^m (Task_j^i, T[t_1^j, t_2^j], C_j^i, QoS_j^i) \quad \text{Où} \quad \dots \dots (5)$$

$$C_j^i = \bigcup_{k=1}^l c_{jk}^i \quad \dots (6) \quad QoS_j^i = \sum_{k=1}^l w_k \quad qos_{jk}^i \quad \dots (7)$$

Rôles : sont les rôles joués par l'agent dans les organisations dont il appartient.

$$Roles = \bigcup_{i=1}^n \left(O_i \left(\bigcup_{j=1}^m role_j^i \right) \right) \quad \dots \dots (8)$$

Pour jouer un rôle l'agent réalise des capacités selon un plan.

$$role_i = \bigcup_{j=1}^n (cp_j, T[t_1^j, t_2^j]) \quad \dots \dots (9)$$

Où $T[t_1^j, t_2^j]$ est l'intervalle de l'exécution de la capacité.

Problème de sélection

Nous voulons composer un nouveau système en utilisant les composants des N systèmes qui sont déjà opérationnels. Supposant que les besoins du nouveau système sont bien définis et modélisés par des agents:

$$Goal = \bigcup_{i=1}^n role_i \quad \dots (10) \quad \text{Où} \quad role_i = \bigcup_{j=1}^{l_i} (cp_j^i, RQoS_j^i) \quad (11)$$

$$capa_{ij} = \begin{cases} QoS_{ij} & \text{Si } cpi \text{ peut être fournie par } aj \text{ avec } QoS_{ij} \\ 0 & \text{Autrement} \end{cases}$$

$$rcapa_{ij} = \begin{cases} 1 & \text{Si } cpi \text{ peut être fournie par } a_j \text{ tq } QoS_{ij} \geq RQoS_i \\ 0 & \text{Autrement} \end{cases}$$

$$roleag_{ij} = \begin{cases} 1 & \text{Si l'agent } j \text{ participe dans le rôle } i \\ 0 & \text{autrement} \end{cases}$$

$$roleag_{ij} = \begin{cases} 1 & \text{Si l'agent } j \text{ participe dans le rôle } i \\ 0 & \text{autrement} \end{cases}$$

$$selag_i = \begin{cases} 1 & \text{si l'agent } i \text{ est c oisi pour former le SdS} \\ 0 & \text{autrement} \end{cases}$$

$$I_{ij} = \begin{cases} ic_{ij} & \text{interaction entre agent } i \text{ et agent } j \text{ avec un coût } ic_{ij} \\ 0 & \text{Autrement} \end{cases}$$

$$colQoS_{ij} = \begin{cases} QoS_{sys} & \text{les agents } i \text{ et } j \text{ appartiennent au même système} \\ QoS_{role} & \text{les agents } i \text{ et } j \text{ participe au même rôle} \\ 0 & \text{Autrement} \end{cases}$$

$$TC = \sum_{i \in I_{Gcap}} \sum_{j \in I_{Gag}} selag_i \quad rcapa_{ij} \quad capa_{ij} + \sum_{i \in I_{Gag}} \sum_{i \in I_{Gag}} selag_i \quad ic_{ij} \quad \dots (12)$$

$$TQoS = \sum_{i \in I_{Gcap}} \sum_{j \in I_{Gag}} selag_i \quad rcapa_{ij} \quad capa_{ij} + \sum_{i \in I_{Gag}} \sum_{i \in I_{Gag}} selag_i \quad colQoS_{ij} \quad \dots (13)$$

L'objectif est de trouver la solution qui minimise le coût total TC et maximiser la QoS totale TQoS qui est un problème multi-objectif que nous avons essayé de le résoudre par les algorithmes génétiques multi-objectifs.

V. PERSPECTIVES

Nous sommes entrain d'implémenter un prototype de plateforme multi-agent pour l'intégration de systèmes de réseaux de capteurs sans fil en utilisant la plateforme Janus.

VI. CONCLUSION

Trois éléments des SdS constituent un défi pour leur modélisation: l'autonomie des CS, l'aspect organisationnel et plusieurs niveaux de représentation possibles. Nous avons profité des caractéristiques des SMA récurifs et utilisé le pour produire un modèle à base d'agents représentant les SdS et facilitant le contrôle de l'introduction dynamique des CS. Nous avons proposé un mécanisme d'intégration de système en se basons sur le problème de sélection. Un prototype d'intégration de réseaux de capteurs sera implémenté sur la plateforme Janus.

REFERENCES

- [1] A. Hamzi ; M. Koudil; J. P. Jamont, M. Occello. "Multi-Agent Architecture for the Design of WSN Applications". *Wireless Sensor Network*, Vol. 5 No. 2, 2013, pp. 14-25. doi: 10.4236/wsn.2013.52003.
- [2] A. Hamzi ; M. Koudil. "Multi-Agent Meta-model for modeling System of Systems". *Journal Simulation: Transactions of the Society for Modeling and Simulation International*. Submitted July 2016.
- [3] A. Hamzi ; M. Koudil. Approaches and Technologies for System of Systems Integration: A review. *International Journal of Systems Engineering*. Submitted March 2017
- [4] A. Hamzi ; M. Koudil. Literature Review on System of Systems Integration Aspects. *ACM Computing Surveys*. Submitted June 2017
- [5] A. Hamzi ; M. Koudil. " Multi-agent system for System of Systems integration, expected date for submission October 2017.

Modèles de Représentation et d'Analyse des Réseaux Sociaux

Narimene Dakiche

Directeurs de thèse : Fatima Benbouzid-Si Tayeb, Yahya Slimani, Karima Benatchba

Équipe : Optimisation

Abstract—Le présent document récapitule le travail déjà réalisé en tant qu'une partie de notre thèse et présente brièvement les parties sur lesquelles nous travaillons actuellement. Notre thèse traite le problème de l'évolution des communautés dans les réseaux sociaux. Dans notre travail, nous nous intéressons essentiellement au problème de suivi des changements liés aux communautés ainsi que la prédiction de leurs futures transformations.

Dans le cadre de notre thèse, nous nous intéressons au problème de détection des communautés dynamiques et particulièrement la détection des changements liés aux communautés extraites d'un réseau social, pour prédire par la suite les éventuelles transformations qu'elles pourraient subir dans le temps (voir Fig.2).

I. INTRODUCTION ET PROBLÉMATIQUE

II. TRAVAUX RÉALISÉS

Les réseaux sociaux peuvent être modélisés par des graphes dont les noeuds sont vus comme des entités en interaction. L'étude de ces interactions permet d'en tirer des informations intéressantes pour diverses applications (marketing, anti-terrorisme, ... etc). Un axe de recherche qui a reçu beaucoup d'intérêt ces derniers temps dans l'analyse des réseaux sociaux (ARS) est la détection des communautés. Une communauté peut être vue comme un ensemble d'entités ayant beaucoup d'interactions entre elles et peu d'interactions avec l'extérieur [1].

A. Synthèse bibliographique

Étant donné que notre thèse comporte essentiellement deux axes: l'évolution des communautés et la prédiction de leurs futures transformations; nous avons commencé par effectuer, dans un premier temps, une analyse pertinente des travaux sur la détection des communautés dynamiques et le suivi de leur évolution dans le temps. A travers cette étude, nous avons établi une classification des travaux existants qui a fait l'objet d'un survey que nous avons soumis dans le journal *Information Processing & Management*.

Dans les approches traditionnelles de l'ARS, la plupart des études réalisées sur la détection des communautés ont été effectuées sur des réseaux statiques, où le comportement des différents individus reste figé à un instant donné. Cependant, dans la vie réelle, la plupart des réseaux sociaux évoluent au fil du temps dû à l'arrivée ou au départ des individus, et à la disparition ou l'apparition de nouvelles relations entre eux: on parle de réseaux sociaux dynamiques.

Nous avons effectué également une synthèse bibliographique des travaux qui traitent la prédiction de l'évolution des communautés. Vu le petit nombre de travaux qui traitent cette problématique, nous avons décidé de commencer par cette partie tout en exploitant des méthodes issues de la littérature dans la phase de détection des changements que pourrait subir une communauté dans le temps.

Les réseaux sociaux dynamique peuvent être modélisés comme une série de réseaux statiques (instantanés), où chaque réseau statique correspond à une instance du réseau social sur une période précise (voir Fig.1). L'évolution des communautés dans ce cas est défini par un ensemble de transformations allant d'un instantané à un autre comme la fusion, la division, la réduction, la formation ...etc [2], [3].

B. Vers une première contribution

Nous avons travaillé au cours de la deuxième année de thèse sur la prédiction de l'évolution des communautés. Nous avons proposé une nouvelle modélisation pour le problème de prédiction dans laquelle nous avons utilisé un ensemble riche de caractéristiques pour décrire les communautés sur chaque instantané. Afin de valider notre approche, nous avons effectué un ensemble de tests sur des réseaux réels de Facebook et DBLP. Cette étude a fait l'objet d'un papier que nous avons soumis à la conférence WSDM 2018.

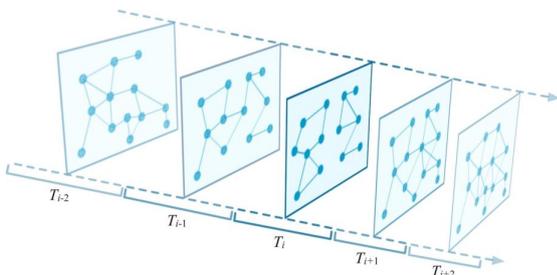


Fig. 1: Un réseau social dynamique.

C. Projet de fin d'étude

Nous avons proposé un projet de fin d'étude pour l'année 2016-2017 à travers lequel nous avons pu réaliser avec les étudiants de cinquième année une plateforme de tests pour la prédiction de l'évolution des communautés dans les réseaux sociaux. La plateforme EPredictor (voir Fig.3) est désormais notre moyen d'effectuer les tests sur la prédiction pour valider nos idées. Elle nous permet de comparer nos résultats avec différentes méthodes de la littérature.

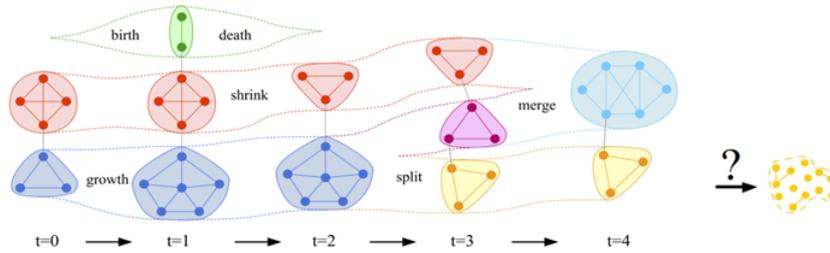


Fig. 2: Suivi de l'évolution des communautés

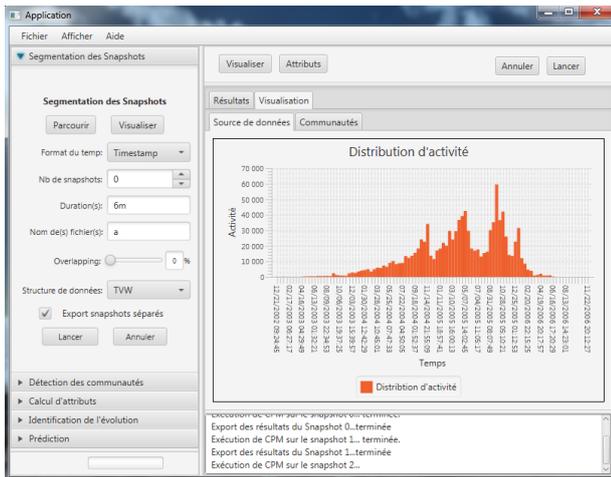


Fig. 3: La plateforme EPredictor

REFERENCES

- [1] M. Girvan and M.E.J. Newman. "Community structure in social and biological networks". *Proceedings of the National Academy of Sciences*, 99(12) :7821-7826, 2002.
- [2] S. Asur and S. Parthasarathy. "An event-based framework for characterizing the evolution of interaction graphs," 2007.
- [3] P. Bródka, P. Kazienko, and B. Kosłoczyk. "Predicting group evolution in the social network," in *International Conference on Social Informatics*, pp. 54-67, Springer, 2012.

III. PERSPECTIVES

Afin de respecter le plan que nous avons établi, avec mes directeurs de thèse, cette troisième année sera consacrée à la proposition d'une nouvelle approche qui permet de détecter les communautés dynamiques et suivre leur évolution dans le temps simultanément. Dans ce sens, nous avons proposé un projet de fin d'étude pour enrichir le champs de l'étude, varier les tests et faciliter la validation en effectuant des comparaisons.

Une fois l'approche validée, nous souhaitons l'intégrer dans notre modèle de prédiction et rassembler les deux propositions dans un article à soumettre dans une revue scientifique à caractère international.

Entre temps, nous planifions de valider notre modèle de prédiction sur un large ensemble de datasets (réseaux sociaux) de différentes nature. De plus, vu la taille énorme des datasets (jusqu'à des millions de noeuds et de liens), nous envisageons d'étudier la possibilité de paralléliser l'exécution de notre modèle.

IV. CONCLUSION

Étant donné que nous n'avons pas encore reçu une réponse favorable de l'une de nos soumissions, il est difficile d'évaluer notre état d'avancement. Cependant, dans l'ensemble, nous estimons que le déroulement des travaux de thèse est globalement positif.

Synergies entre optimisation combinatoire et data mining

SADEG Souhila

Directeurs de thèse : Pr.BENATCHBA Karima, Pr.HABBAS Zineb

Abstract—Notre sujet de thèse traite des synergies existant entre deux disciplines qui sont l'optimisation combinatoire et le data mining en étudiant l'apport de chaque domaine à l'autre. Nos travaux s'articulent autour de deux volets. Le premier concerne l'utilisation des méthodes de résolutions de problèmes d'optimisation combinatoire, et plus particulièrement les métaheuristiques, pour résoudre des problèmes de data mining. Quant au deuxième volet, il s'intéresse à l'utilisation des techniques de data mining pour améliorer les performances des métaheuristiques.

I. INTRODUCTION

Le data mining (DM) est une discipline qui est apparue au milieu des années 1990. Elle est souvent décrite comme étant le coeur du processus de découverte (extraction) de connaissances KDD (Knowledge Discovery and Data mining). Son objectif est d'extraire et d'exploiter des connaissances potentiellement utiles et souvent cachées dans de grands volumes de données. Ce domaine connaît ces dernières années un engouement de la part de chercheurs de différents horizons et ce pour deux raisons : 1) la constante croissance du volume de données à traiter, 2) la qualité de l'interprétation que permettent les analyses approfondies qu'offre le data mining en comparaison avec de simples résultats statistiques. Le Data mining est constitué de quatre tâches principales [1]: 1) la classification supervisée qui vise à expliquer ou classer des caractéristiques d'un individu donné suivant les valeurs prises par le reste des caractéristiques; 2) la classification non supervisée (aussi appelée clustering) qui a pour objectif de construire des groupes homogènes d'individus. Elle exprime les similarités découvertes à l'intérieur des groupes et les dissimilarités entre les groupes; 3) les règles d'association dont l'objectif est de découvrir des similitudes entre les individus; 4) la sélection d'attributs, qui peut être appliquée avant l'une des trois tâches précédentes afin de pallier le problème du grand nombre d'attributs en sélectionnant les plus pertinents d'entre eux. Les problèmes posés par le data mining constituent des défis importants auxquels s'attaquent plusieurs domaines de recherche tels que les statistiques, la théorie de l'information, les bases de données, l'apprentissage automatique, la recherche opérationnelle [1] et le domaine de l'optimisation plus particulièrement. En effet, beaucoup de travaux utilisant des méthodes de résolutions de problèmes d'optimisation combinatoire, et plus particulièrement les métaheuristiques, pour résoudre des problèmes de data mining ont été publiés [3], [5], [6], [7], [8]. Ces dernières années ont vu l'émergence d'une nouvelle tendance qui consiste à utiliser des techniques du data-mining pour extraire de la connaissance pouvant aider les métaheuristiques à résoudre des problèmes NP-Difficiles [2]. Cette association est désignée par le mot "synergies"¹ pour

¹Mise en commun de plusieurs actions concourant à un effet unique et aboutissant à une économie de moyens

exprimer la combinaison de méthodes issues des deux disciplines pour la résolution de problèmes [1], [4]. Ce document présente l'état d'avancement de nos travaux de thèse. Nous commençons par rappeler notre problématique de recherche puis nous présentons les travaux réalisés à ce jour. Enfin, nous terminons ce rapport par des perspectives pour l'année en cours.

II. PROBLÉMATIQUE

Notre thèse porte sur l'étude des synergies entre le domaine de l'optimisation combinatoire et le data mining. Ces travaux s'articulent autour de la coopération entre ces deux domaines. Dans une première partie de la thèse, une étude est menée sur l'apport de l'optimisation combinatoire au Data Mining. Cette étude se focalise sur l'utilisation des métaheuristiques dans les différentes tâches du data mining. En particulier, nous utilisons la métaheuristique Bee swarm Optimization (BSO) pour la sélection d'attributs. La deuxième partie de la thèse est consacrée à la façon dont les techniques de data mining peuvent être utilisées pour améliorer les méthodes de résolution de problèmes d'optimisation comme les métaheuristiques. Nos travaux sont axés sur ce volet qui est le plus récent.

III. TRAVAUX RÉALISÉS

Cette thèse étant composée de deux volets, nous présentons dans cette section nos travaux selon le volet dans lequel ils s'inscrivent tel que l'illustre la figure 1 qui indique l'intervalle de temps consacré à chaque travail.

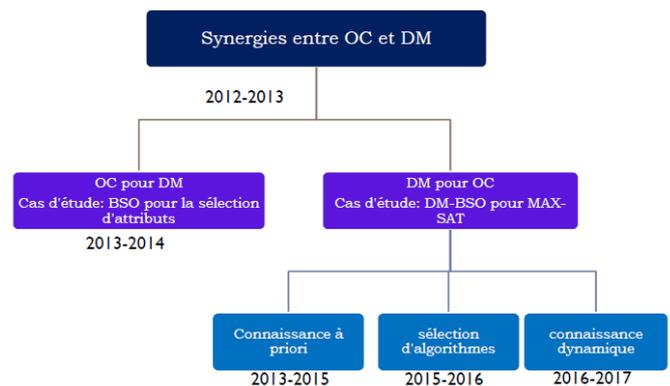


Fig. 1. plan des travaux de thèse

A. Métaheuristiques pour le data mining

Nous nous sommes intéressés à la sélection d'attributs qui comme beaucoup de tâches du data mining peut être formulée comme étant un problème d'optimisation auquel les

métaheuristiques sont souvent appliquées. Dans notre cas, la métaheuristique BSO a été adaptée au problème et testée sur des benchmarks très utilisés dans les travaux de Data mining. un article intitulé "BSO-FS: Bee Swarm Optimization for feature selection in Classification" a été présenté dans "13th International Work-Conference on Artificial Neural Network" (IWANN'2015) et publié dans *Advances in Computational Intelligence* (Springer) .

B. Data mining pour les métaheuristiques

Nos travaux dans ce volet étudient l'utilisation du data mining pour booster les performances des métaheuristiques avec comme cas d'étude la métaheuristique BSO appliquée au problème MaxSAT. Le data mining est utilisé dans le but d'extraire de la connaissance pouvant servir à améliorer les résultats en termes de qualité et/ou en termes de temps d'exécution. Cette connaissance peut être de deux types selon le moment où elle est extraite. Elle est dite "à priori" si elle est extraite avant le processus de résolution et "dynamique" si son extraction se fait durant ce processus. [2] [1].

- *Connaissance à priori*. Nous visons avec ce type de connaissance à extraire de la connaissance à partir d'une instance du problème avant sa résolution. Deux travaux ont été réalisés, basés sur l'extraction de la connaissance à priori en utilisant des techniques de data mining.
 - Le premier travail intègre la sélection d'attributs à BSO en vue de réduire la taille de l'instance dans l'objectif de réduire le temps de calcul. Pour cela nous commençons par représenter l'instance sous forme d'un dataset. Plusieurs représentations ont été proposées en utilisant deux types d'organisations, deux types de codages, et deux approches d'évaluation des attributs ce qui a donné lieu à huit variantes de l'algorithme hybride nommé FS-BSO. Elles ont toutes été implémentées et testées sur différentes instances de MaxSAT.
 - Le second travail a pour but d'utiliser le data mining dans le but de sélectionner pour une instance MaxSAT donnée, l'algorithme qui lui est le plus approprié. En effet, il est connu qu'un algorithme ne peut pas être meilleur que les autres sur toutes les instances de tous les problèmes. Aussi, face à une instance donnée, la question du choix de l'algorithme le plus adapté se pose. Une manière d'y répondre consiste à utiliser l'apprentissage automatique en exploitant les connaissances sur les performances d'un ensemble d'algorithmes sur des instances résolues précédemment. Ainsi, en présence d'une nouvelle instance, le choix se portera sur l'algorithme ayant donné les meilleurs résultats sur des instances qui lui sont similaires. Afin d'évaluer cette similarité, les instances sont considérées comme étant des objets ayant des caractéristiques. Celles-ci doivent être choisies par des experts maîtrisant la nature du problème en question. Dans notre travail, un ensemble de 5 algorithmes a été utilisé sur un ensemble 1400 instances.
- *connaissance dynamique*. Ce type de connaissance est extrait pendant le processus de recherche en exploitant les résultats intermédiaires. Dans nos travaux nous proposons d'intégrer des techniques du data mining à BSO afin d'accélérer le processus d'évaluation. Cinq algorithmes hybrides sont proposés:
 - CLUS-BSO: cet algorithme utilise la technique de clustering pour regrouper les solutions et désigner un représentant à chaque cluster. Seul les représentants sont évalués tandis que les autres solutions appartenant au même groupe sont considérées de qualité similaire.
 - FIM-BSO: dans cet algorithme la technique de recherche des items set fréquents (Frequent Item Set) est appliquée sur l'ensemble des meilleures solutions obtenues après un certain nombre d'itérations de BSO afin d'extraire les caractéristiques qui leur sont communes. Ceci permet, au cours des itérations suivantes de n'évaluer que les solutions présentant ces caractéristiques.
 - CLAS-BSO: dans cet algorithme les solutions sont classées selon leur qualité en deux classes "bonne" ou "mauvaise". Un modèle est ensuite construit pour classer les solutions suivantes afin que seuls les solutions estimées bonnes soient évaluées.
 - SSC-BSO: dans cet algorithme nous utilisons une classification semi-supervisée qui tolère l'absence d'étiquette pour une partie des données ce qui réduit le nombre d'évaluation nécessaires avant la construction du modèle.
 - REG-BSO: la régression est une méthode utilisée pour prédire la valeur d'un attribut numérique en utilisant les valeurs des autres attributs. Dans cet algorithme la valeur à prédire est la fitness d'une solution. le modèle construit évite d'avoir recours à l'évaluation complète.

IV. CONCLUSION

Ce présent rapport présente l'état d'avancement de nos travaux de thèse. Après un rappel du sujet et de la problématique traitée, nous avons présenté brièvement les travaux réalisés. Un article présentant nos résultats en utilisant la connaissance à priori a été soumis en octobre 2016 à la revue *Swarm and Evolutionary Computation* (revue de rang A*) qui l'a rejeté après deux rounds de reviewing. Cet article a été corrigé suivant les remarques des reviewers et soumis à la revue *Neurocomputing* (de rang A) le 03/09/2017. L'année en cours sera consacrée à la rédaction d'autres articles traitant des travaux utilisant le data mining pour l'extraction de la connaissance dynamique et la sélection d'algorithmes.

REFERENCES

- [1] Corne, D., Dhaenens, C., Jourdan, L. 2012. Synergies between operations research and data mining: The emerging use of multi-objective approaches. *European Journal of Operational Research*, 221(3), 469-479.
- [2] Jourdan, L., Dhaenens, C., Talbi, E.-G., 2006. Using Data Mining Techniques to Help Metaheuristics : A Short Survey. In: LNCS (Ed.), *Hybrid Metaheuristics*. Vol. 4030. Springer, pp. 57- 69.

- [3] Karasozen, B., Rubinov, A., Weber, G.-W., 2006. Optimization in data mining. *European Journal of Operational Research* 173 (3), 701 ? 704.
- [4] Meisel, S., Mattfeld, D., 2010. Synergies of operations research and datamining. *European Journal of Operational Research* 206 (1), 1 ? 10.
- [5] Olafsson, S., 2006. Introduction to operations research and data mining. *Computers and Operations Research* 33 (11), 3067 ? 3069.
- [6] Olafsson, S., Li, X., Wu, S., 2008. Operations research and data mining. *European Journal of Operational Research* 187 (3), 1429 ? 1448.
- [7] Dudas, S., FrantzÃon, M., Ng, AHC., 2011. A synergy of multi-objective optimization and data mining for the analysis of a flexible flow shop. *Robotics and Computer-Integrated Manufacturing*, Volume 27, Pages 687-695
- [8] Yeh, W-C., 2012. Novel swarm optimization for mining classification rules on thyroid gland data , *Information Sciences*, Volume 197, Pages 65-76
- [9] Sadeg, S., Hamdad, L., Benatchba, K., Habbas, Z., 2015. BSO-FS: Bee Swarm Optimization for Feature Selection in Classification. In *Advances in Computational Intelligence* (pp. 387-399). Springer International Publishing.

Incorporation des données non-relationnels dans les systèmes d'information décisionnel

Mohamed Lamine CHOUDER

Directeurs de thèse : Pr. Rachid CHALAL

Équipe : MSI

Résumé—Au cœur du système d'information décisionnel, l'entrepôt de données intègre et stocke des données, issue de plusieurs sources, permettant leur exploitation moyennant l'analyse multidimensionnelle et le traitement analytique en ligne. Etant donné la diversité des sources de données relationnels et non-relationnels, notre objectif de la thèse est de proposer une approche pour supporter l'analyse multidimensionnelle et le traitement analytique en ligne sur les sources de données non-relationnels. Ce document résume l'état d'avancement de notre travail de recherche.

I. INTRODUCTION

Dans l'aide à la décision stratégique, les décideurs ont besoin d'exploiter les informations (internes) provenant des sources opérationnelles ainsi que les informations (externes) émanant de l'environnement de l'entreprise. Au cœur du système d'information décisionnel, l'entrepôt de données (ED) intègre et stocke des données, issue de plusieurs sources, dans une forme multidimensionnelle permettant leur exploitation moyennant plusieurs outils d'analyse à savoir le traitement analytique en ligne.

Depuis quelques années, les entreprises ont adopté une nouvelle génération de bases de données (BD) dites non-relationnels pour stocker les grandes masses de données générées par les applications modernes. En effet, les BDs non-relationnels ont émergé comme une alternative des BDs relationnels dans plusieurs implémentations à savoir les applications Big Data, Web temps réel et Internet des objets. A cet effet, l'utilisation des BDs non-relationnels a produit des grandes masses de données contenant des informations précieuses pouvant être profitablement intégrées dans le système d'information décisionnel de l'entreprise. Notre intérêt dans la thèse est porté sur l'incorporation des données non-relationnels dans les systèmes d'information décisionnel.

II. PROBLEMATIQUE

Plusieurs approches de conception multidimensionnelle d'EDs ont été proposé dans la littérature, qui peuvent être classé en deux catégories : orienté-données et orienté-besoin. Les approches orienté-données défini le schéma multidimensionnel en analysant les sources ; Alors que, les approches orienté-besoin analyse les besoins des utilisateurs pour définir le schéma.

D'après notre étude des approches orienté-besoin, nous avons constaté qu'ils évoquent sommairement l'intégration des informations externes issue majoritairement des sources de données non-relationnels. Malheureusement, la nature de ces sources rend la conception d'un schéma multidimensionnel

qui satisfait les besoins des décideurs, une tâche complexe et difficile.

Dans une deuxième étape, nous nous sommes intéressés par les approches orienté-données pour la conception multidimensionnelle à partir des sources non-relationnels. La majorité des sources traitées dans cette catégorie contient des données de nature externe, comme les données graph (réseaux sociaux), données ouvertes (comme RDF) et données XML. Malheureusement, les approches existantes ne sont pas adaptées aux BDs non-relationnels.

III. TRAVAUX REALISES

Dans cette section nous présentons le périmètre de la thèse ainsi que les travaux qui ont été publiés ou soumis à des conférences ou à des journaux.

A. Première proposition

Durant les premières années de la thèse, nous nous sommes focalisés principalement sur les approches de conception multidimensionnelle orienté-besoin et plus particulièrement orienté-but pour plusieurs raisons : (i) ces approches fournissent une meilleure définition des besoins ; (ii) Les objectifs de l'organisation au niveau stratégique sont considérés comme la ressource principale pour identifier les besoins en matière d'information décisionnel. Notre étude a conclu que ces approches focalisent sur les informations issues des sources de données relationnels.

Notre objectif était de proposer une approche qui permet de spécifier les besoins en matière d'information externe provenant des sources non-relationnels. Afin d'exploiter les travaux faites par notre équipe sur la modélisation de l'environnement, nous avons opté pour une approche basée sur les modèles. Notre première contribution a porté sur l'extension du modèle des buts par les besoins en informations externes. Notre deuxième contribution porte sur un ensemble de guidelines proposées pour utiliser les modèles support : Le premier a pour but la définition des besoins en informations externe alors que le deuxième représente une structure de l'information se rapportant sur l'environnement. Cette approche a été évaluée par une étude de cas.

Notre première proposition a donné lieu deux communications [2,3] à la conférence IFIP CIIA'2015 (Springer). La première a porté sur notre approche et la deuxième sur une étude comparative des méthodes orienté-besoin de conception multidimensionnelle.

A. Deuxième proposition

Durant les deux dernières années de la thèse, nous avons étudié les approches de conception multidimensionnelle à partir des sources de données non-relationnelles. Cela nous a permis de proposer une approche interactive pour supporter l'analyse multidimensionnelle et le traitement analytique en ligne sur les BDs non-relationnelles spécifiquement orienté-document. Pour ce faire, la première étape consiste à l'extraction des structures à partir des collections de données semi-structurées, stockées dans ces bases et caractérisées par l'absence de schéma. Par la suite, ces structures sont utilisées pour définir un schéma multidimensionnel qui va permettre aux utilisateurs d'exécuter une session analytique commençant par la formulation de requêtes multidimensionnelles suivi par des opérations analytiques. Ceci est achevée par la découverte automatique des hiérarchies multidimensionnelles en se basant sur la fouille des dépendances fonctionnelles à partir des données ; afin d'assurer une meilleure performance, les hiérarchies sont construites de façon incrémentiel à la base de la session courante de l'utilisateur. L'approche a été implémenté dans MongoDB, l'un des plus connus système de stockage de BDs non-relationnelles. L'approche est évaluée par des expérimentations sur des données réels pour prouver sa compatibilité avec une utilisation en temps réel.

Cette deuxième proposition a donné lieu à une communication [3] à la conférence DOLAP'2017(EDBT/ECDT workshops) et un article de journal soumis et accepté sous réserve d'une correction mineure [4].

IV. PERSPECTIVES

Dans le mois à venir, nous allons corriger et soumettre l'article de journal. Puis, nous allons entamer la rédaction de la thèse.

V. CONCLUSION

Dans ce document nous avons présenté une brève description des travaux effectués dans notre thèse. Nous estimons que l'avancement dans la thèse est globalement positif.

REFERENCES

- [1] M. L. Chouder, R. Chalal, and W. Setra, "Requirement analysis in data warehouses to support external information," in *CIIA'2015*, 2015.
- [2] W. Setra, M. L. Chouder, and R. Chalal, "Engineering the Requirements of Data Warahouses: a Comparative Study of Goal-Oriented Approaches," in *CIIA'2015*, 2015.
- [3] M. L. Chouder, S. Rizzi, R. Chalal, Enabling Self-Service BI on Document Stores. In *EDBT/ECDT Workshops*, 2017.
- [4] M. L. Chouder, S. Rizzi, R. Chalal. Exodus: Exploratory OLAP over Document Stores. soumis à *Information Systems- Elsevier*, accepté avec une correction mineure.

La reconnaissance de l’empreinte palmaire latente

Touka Faisal

Directeurs de thèse : Pr. Karima Benatchba, Pr. Mouloud Koudil

Équipe : Biométrie

Résumé— La reconnaissance des empreintes palmaires latentes présente un déficit pour les chercheurs du fait que qu’elles soient d’une qualité dégradée, ne couvrent qu’une partie de la surface de la paume de la main, de plus qu’elles possèdent un background complexe. Dans notre thèse nous visons développer une approche de reconnaissance d’empreinte palmaire latente en s’inspirant d’autre domaine tel que la reconnaissance digitale, le recalage d’image ainsi que les techniques d’optimisation.

I. INTRODUCTION

Depuis l’aube des temps, la reconnaissance des individus a pris une importance dans la vie humaine. Son intérêt a évolué de la simple association des personnes aux événements ou aux transactions dans un cadre social ou commercial, à l’identification à grande échelle dans le cadre de la sécurité des nations. L’une des premières applications, dans un cadre officiel du concept de la reconnaissance biométrique, fut dans le domaine criminalistique. Vucetich établit la première identification d’un criminel par sa trace d’empreinte digitale en 1892.

Depuis ce temps, l’utilisation des empreintes digitales est devenue décisive dans l’identification des personnes impliquées dans des infractions pénales par les experts de la criminalistique. Cette utilisation de données biométriques ne s’est pas limitée aux empreintes digitales. En 1986, et pour la première fois, l’ADN va confondre l’auteur d’un double meurtre dans ce qu’on appelle l’affaire « Colin Pitchfork » (Champod, 2009). D’autres preuves biométriques deviennent de plus en plus répandues : l’empreinte palmaire, les informations fournies par les enregistrements vidéos de surveillance tels que les images de visages, la démarche..

Beaucoup de chercheurs se sont intéressés à la reconnaissance des empreintes digitales preuve en est de la riche littérature qui existe. Cependant, bien que les spécialistes assurent que 30% des évidences collectées sur les scènes de crime sont des empreintes palmaires latentes [8], peu de chercheurs se sont penchés sur ce problème. Ce sujet s’inscrit dans ce contexte.

II. PROBLEMATIQUE

La reconnaissance d’empreinte palmaire dans le cas usuel, civil ou commercial, n’est autre qu’un processus de comparaison de deux images d’empreintes complètes de qualité contrôlée. Cependant, dans le cas d’enquête criminelle, les marques d’empreinte palmaire collectées, dite empreintes latentes, sont partielles, de mauvaise qualité et sans forme structurelle définies. On parle ainsi de comparaison d’empreinte latente aux empreintes complètes. Cette différence d’objectifs et de données a rendu difficile l’adaptation des approches conçues pour le cas usuel d’empreinte complète et a obligé les chercheurs à exploiter d’autres pistes.

D’autre part, la nature de l’empreinte palmaire similaire à celle digitale a incité les chercheurs à exploiter les concepts et les approches conçues pour la reconnaissance digitale. A titres

d’exemple, des points caractéristiques robustes communs, telles que les minuties, ayant déjà prouvés leurs efficacités ont été utilisées pour identifier les empreintes palmaires latentes. L’utilisation des minuties ne présente pas la seule stratégie adaptable dans notre cas. Plusieurs autres notions déjà appliquées dans la reconnaissance digitale peuvent être exploitées. Malgré cette variété apparente de choix et d’idées, la reconnaissance d’empreintes palmaires latentes reste peu abordée dans les travaux de recherche.

Notre sujet consiste à concevoir une approche de reconnaissance d’empreinte palmaire latente. Cette approche devra toucher aux différentes étapes du processus de reconnaissance à travers :

- Le choix des caractéristiques à exploiter ainsi que leur représentation dans la phase d’extraction des caractéristiques.
- Le choix des techniques de comparaison des différentes caractéristiques et de calcul des scores dans la phase d’appariement.
- Le choix d’une stratégie de fusion des scores

III. TRAVAUX REALISES

Nous avons initié notre recherche par l’étude du contexte général de notre sujet tel que : des généralités sur la reconnaissance biométrique et ses cas d’utilisation dans le domaine criminalistique, des généralités sur l’empreinte palmaire, ses caractéristiques biométriques, ses types de reconnaissance et les approches de reconnaissance dans le cas usuel et commercial ainsi que des généralités sur la reconnaissance digitale et ses différentes approches.

Vu leurs nombres limitées, nous avons pu élaborer par la suite une étude détaillée et approfondie des différents travaux touchants de près ou de loin la reconnaissance d’empreintes palmaires qu’elles soient latentes, partielle ou même complète de haute résolution tel que : les travaux de l’équipe de Jain (l’université de Michigan) [1][2][3][4], les travaux de l’équipe de Li (l’université de Nankai en Chine) [5][6][7], les travaux de Laadjel (Queen’s University Belfast UK) [8][9], les travaux de l’équipe de Wang (L’université de Madrid) [10][11], Les travaux de l’équipe de Dai (l’université Tsinghua, chine) [12][13], le travail de l’équipe de Cappelli (l’université de Bologne, Italie)[14], Les travaux de l’équipe de Singh (l’université de Lisbonne) [15][16] et dernièrement le premier travail de conception d’une base de donnée publique dédiée à cette problématique [17]

Cette étude nous a permis de tracer une comparaison détaillée entre les différents travaux du domaine selon plusieurs volets : la publication (auteurs, date, type), les techniques utilisées (Inspiration du travail, Classification de l’appariement principal, Fusion, Amélioration de qualité, Calcul du champ d’orientation, Caractéristiques extraites, Elimination des minuties..) ainsi que les BDD et les tests

De même nous avons pu effectuer une classification des approches étudiées; et schématiser le processus général

complet d'une reconnaissance type d'empreinte palmaire latente.

Cette étude bibliographique comparative fait le sujet d'un article de revue en cours de rédaction.

Après l'étude des différents travaux dans le domaine et en constatant l'apport que peut apporter l'application des techniques utilisées dans d'autres domaines nous avons conçu notre approche de reconnaissance où nous avons choisi les minuties comme caractéristiques extraites, les informations globales de ces derniers (emplacements, orientations) comme représentation des caractéristiques extraites et l'appariement global comme stratégie d'appariement. Nous avons formalisé le problème d'appariement des minuties comme un problème de « Point Pattern Matching » qui est un problème très connu dans la littérature et utilisé dans le domaine de recalage d'image et peut être résolu comme un problème d'optimisation. Nous avons choisi les algorithmes génétiques (AG) pour résoudre ce problème et ceci après avoir étudié différents articles d'application des AG pour résoudre le PPM (ex : [18]) ou pour la reconnaissance digitale (ex : [19]).

Nous avons implémenté et tester le processus de prétraitement et d'extraction de minuties en utilisant des techniques classiques utilisée dans la reconnaissance digitale. Plusieurs fonctions objectives estimant la ressemblance entre deux vecteurs de minuties ont été étudiées, implémentées, testées et comparées et une nouvelle proposition montrant son efficacité sur la base de [18] a été mise en place. Un algorithme génétique est implémenté proposant une nouvelle méthode de croisement à base de la moyenne et la moyenne pondérée et de remplacement à base du tournoi. Cet algorithme est en phase de test moyennant le cluster IbnBadis

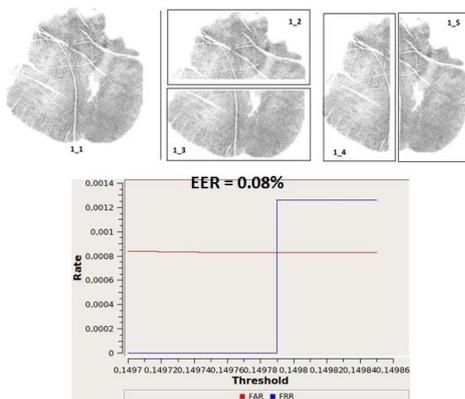


Figure 1- Resultat de la reconnaissance sur une partie de la BDD

IV. PERSPECTIVES

Nous avons comme objectifs

- La finalisation de l'article sur la première variante de notre approche de reconnaissance à base d'appariement des minuties par les AG (parcours des transformations).
- L'amélioration de cette variante à travers l'amélioration de la phase de prétraitement.
- L'implémentation d'une deuxième variante de notre approche de reconnaissance (parcours des fonctions map).
- Le suivi d'un PFE qui implémente une autre approche de reconnaissance à base des descripteurs locaux.

V. CONCLUSION

Le développement informatique a ouvert les portes vers l'automatisation de plusieurs processus qui ont resté manuels

jusqu'au là. Parmi les nouvelles pistes de recherche nous trouvons la reconnaissance automatique des empreintes palmaires latentes. Cette dernière fut le sujet de plusieurs travaux de recherche récents mais qui reste peu nombreux.

Comme tout nouveau domaine de recherche, il se trouve intéressant de profiter des avancements faits dans des domaines proches. Nous avons ainsi choisi de s'inspirer de la reconnaissance digitale pour l'utilisation des minuties, du « point pattern matching » pour formaliser notre problème comme un problème d'optimisation et des heuristiques d'optimisation évolutive (les AG) pour sa résolution.

VI. REFERENCES

- [1] D. Nicolae, J. K. Anil and M. Kanti, "Matching of palmprints", Pattern Recognition Letters.-Elsevier, 2002, Vol.23, pp.477-485.. (references)
- [2] J. K. Anil and D. Meltem, "On Latent Palmprint Matching" Department of Computer Science & Engineering, 3115 Engineering Building, Michigan State University, East Lansing, MI, 2008. (references)
- [3] J. K. Anil and F. Jianjiang, "Latent Palmprint Matching", IEEE Transactions on pattern analysis and machine intelligence, Juin 2009, 6 : Vol. 31, pp. 1032-1047. (references)
- [4] L. Eryun, J. K. Anil and T. Jie, "A Coarse to Fine Minutiae-Based Latent Palmprint Matching", IEEE transactions on pattern analysis and machine intelligence. , 2013. 10 : Vol. 35. , pp. 2307-2322 (references)
- [5] L. Jijy [et al.], "A new point pattern matching method for palmprint", International Conference on Artificial Intelligence and Pattern Recognition (AIPR07), 2007. (references)
- [6] L. Jijy [et al.], "A Research on Offline Palmprint Identification" ,World Congress on Computer Science and Information Engineering, 2009. (references)
- [7] T. Zechao [et al.], "Minutiae-Based Offline Palmprint Identification System" ,Global Congress on Intelligent Systems, 2009. (references)
- [8] L. Moussadek [et al.], "Degraded partial palmprint recognition for forensic investigations",16th IEEE international conference on Image processing, 2009, pp. 1497-1500. (references)
- [9] L. Moussadek [et al.], "Partial Palmprint Matching Using Invariant Local Minutiae Descriptors", Transactions on data hiding and multimedia security V, Springer, 2010, Vol. 5 (references)
- [10] W. Ruifang, R. Daniel and F. Julian, "Latent-to-full palmprint comparison based on radial triangulation under forensic conditions" Biometrics (IJCB), 2011 International Joint Conference on., 2011. - pp. 1-6. (references)
- [11] W. Ruifang, R. Daniel and F. Julian, "Improving radial triangulation-based forensic palmprint recognition according to point pattern comparison by relaxation",In Biometrics (ICB), 2012 5th IAPR International Conference on.,New Delhi, pp. 427-432. . (references)
- [12] D. Jifeng and Z. Jie, "Multifeature-Based High-Resolution Palmprint Recognition", Transactions on Pattern Analysis and Machine Intelligence, Mai 2011, 5 : Vol. 33. - pp. 945-957. (references)
- [13] D. Jifeng, F. Jianjiang and Z. Jie, "Robust and efficient ridge-based palmprint matching", Pattern Analysis and Machine Intelligence, IEEE Transactions on., August 2012, 8 : Vol. 34, pp. 1618-1632 . (references)
- [14] C. Raffaele, FerraraF. Matteo and M. Dario, "A fast and accurate palmprint recognition system based on minutiae", Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on., June 2012, 3 : Vol. 42, pp. 956-962. (references)
- [15] S. Sanchit [et al.], "PP-RIDER: A Rotation-Invariant degraded partial palmprint recognition technique", Signal Processing Conference (EUSIPCO), 2012, pp. 1499 – 1503 (references)
- [16] S. Sanchit, C. P. Lobato and S. L. Ducla, "Improved rotation-invariant degraded partial palmprint recognition technique", In Biometrics and Forensics (IWBF), 2013, Lisbon , pp. 1-4. (references)
- A. Morales, "LPIDB v1. 0-Latent palmprint identification database", (IJCB), IEEE International Joint Conference on, 1-6, 2014. (references)
- [18] Tico, M., & Rusu, C. (1998, September). Point pattern matching using a genetic algorithm and Voronoi tessellation. In Signal Processing Conference (EUSIPCO 1998), 9th European (pp. 1-4). IEEE. (references)
- [19] W.Sheng [et al.] "A memetic fingerprint matching algorithm". Information Forensics and Security, 2007 IEEE Transactionson, 2(3),402-412(references)

Systemes de contraintes géométriques : décomposition et résolution

ABID khaled

Directeurs de thèse :AIT-AOUDIA Samy Co-directeurs de thèse :Dominique Michelucci

Résumé—Cette année, nous nous sommes mis comme objectif de faire une étude critiques sur les différentes méthodes de résolution des contraintes géométriques bien contraints quand les inconnus sont de grande taille.

I. INTRODUCTION

Les contraintes géométriques sont au cœur des applications du CAO et DAO, et entre aussi dans plusieurs contextes de modélisation géométrique tels que : robotique, modélisation moléculaire, apprentissage géométrique... etc.

Elles sont devenues nécessaires après les limites majeures des représentations impératives qui modélisent l'objet par toutes ces données explicites.

Les objets géométriques que veut modéliser l'utilisateur, doivent vérifier certaines propriétés, traditionnellement appelées contraintes. Les contraintes dans les modeleurs classiques n'avaient pas de représentation informatique et c'était à l'utilisateur de les gérer "manuellement" et d'assurer la "cohérence" en cas de modification.

La modélisation par contraintes consiste à vérifier un ensemble de relations entre les différentes parties d'un objet géométrique tels que : points, lignes, cercles et coniques en 2D, ou bien points, lignes, cylindres et sphères en 3D. Les contraintes sont spatiales et incluent des contraintes logiques tel que : incidence, parallélisme, tangences, et des contraintes métriques tel que : distances, angles, et rayon.

II. PROBLEMATIQUE

Actuellement, il y'a plusieurs méthodes de résolution des contraintes géométriques qui réduisent des problèmes de grosse taille et qui peuvent être utilisées interactivement par les dessinateurs comme un outil de conception pendant le processus du dessin.

On se propose dans le cadre de cette thèse de doctorat de faire une étude critique des différentes méthodes proposées et faire ressortir les paramètres pertinents pour le choix d'un modeleur. Les pistes de résolution par équations boites-noires et re-paramétrage seront explorées. Une comparaison, selon des critères pertinents, avec des méthodes proposées sera également menée.

Le travail que nous devons mener dans le cadre de cette thèse de Doctorat est prévu selon l'échéancier global suivant :

I. Faire une recherche bibliographique et un état de l'art du domaine de la modélisation géométrique par contrainte (1 an).

II. Etude Critique des différentes méthodes proposées. Faire ressortir les paramètres pertinents pour le choix d'un modeleur spécifique pour la résolution des contraintes géométriques (1 an).

III. Proposition d'une nouvelle méthodologie de résolution se basant sur le re-paramétrage et la théorie des graphes (1 an).

IV. Exploitation des résultats obtenus (comparaisons et conclusions), publication dans une revue, et rédaction de la thèse (1 an).

III. TRAVAUX RÉALISÉS

Depuis notre première inscription en mois d'octobre 2014, nous avons pu avancer dans la partie de l'état de l'art (Recherche bibliographique) où nous avons étudié quelques thèses et articles, les résumés et faire une synthèse (points forts et points faibles) pour chaque travail.

Nous avons également fait la récolte des données nécessaires pour l'évaluation de notre future travail, où nous avons étudié les approches de modélisation géométriques (Approches algébriques et approches géométriques).

Cette année, nous avons bénéficié d'un stage PNE d'une durée de 08 mois où nous avons profité de faire une étude critique sur les différentes méthodes de résolution des contraintes géométriques bien contraintes.

Parmi les exemples qu'on a implémenté:

- Placement des cercles dans un triangle donné de tel sorte qu'ils soient tangents entre eux et avec les arêtes du triangle. Le principe est d'augmenter le nombre de cercles à chaque fois et d'évaluer le temps de convergence par rapport à chaque méthode d'optimisation.
- Calcul du point le plus proche à un point donné dans un objet.
- Calcul des deux points les plus proches entre deux objets.
- etc.

D'abord, on a utilisé le solveur "Baron" qui est un système de calcul pour résoudre les problèmes d'optimisation non convexe à l'optimisation globale sur Matlab; Malheureusement, on n'a pas pu dépasser des équations avec 10 inconnus (la version payante n'est pas disponible).

Ensuite, on a programmé les mêmes problématiques en appliquant plusieurs algorithmes de résolution d'équations linéaires tel que:

- CG (gradient conjugué)

- Newton CG (newton gradient conjugué)
 - BFGS (Broyden–Fletcher–Goldfarb–Shanno)
 - SLSQP (*Sequential (least-squares) quadratic programming*)
 - Nelder-mead
 - Powell
 - Etc.
- Se donner un maillage triangulaire initial d'une surface de subdivisions ou un profil/contour d'extrusion et modifier les sommets initiaux pour que ça passe par des points donnés (A applique sur blender).
 - Est ce que une méthode jacobian free telle que Jaya ne devient pas meilleure que Newton quand le nb d'inconnues est très grand.
 - Comment résoudre un système d'équations sous-jacent à un système de contraintes: Krylov. ...

Pour cela on a choisi Python comme langage de programmation avec la bibliothèque SciPy qui contient des modules pour l'optimisation, l'algèbre linéaire, les statistiques, le traitement du signal ou encore le traitement d'images.

IV. PERSPECTIVES

Dans l'année qui suit Plusieurs pistes seront explorées pour la réalisation d'un article. Parmi ces piste on sites:

V. CONCLUSION

Pendant cette année, on a pu mettre la main sur les différentes méthodes de résolution de contraintes géométriques, cela nous a permis d'avoir une idée sur les points forts et faibles de chaque méthode quand les inconnus sont de grande taille.

Energy efficiency in new generation networks

Rachid SELT

Directeurs de thèse : Professeur Yacine CHALLAL , Abdelmalik BACHIR

Équipe : LMCS

Abstract—In WSNs, sensor nodes are widely deployed in different environments with different temperature. In consequence, temperature changes affect the quality of the link between a transmitter and a receiver. Higher temperatures reduce the quality of the wireless link by affecting both the receiver and the transmitter. We formulate the energy problem as a linear programming problem, where the objective is to maximize the network lifetime, which is equivalent to the time until the first node is drained of its energy.

I. INTRODUCTION

A wireless sensor network (WSN) [1] is a network formed by a large number of sensor nodes where each node is equipped with a sensor to detect physical phenomena such as light, heat, pressure, etc. WSN's feature easier deployment and better flexibility of devices. With the rapid technological development of sensors, WSN's will become the key technology for IoT.

One of the major causes of energy loss [2] in the WSN node is the idle mode consumption. There is also an energy loss due to packet collision. A third cause of energy loss is the reception of packets not addressed to the node. The fourth major source of wasted energy is the transmission - and possible retransmission- of control packets. Recently, several studies [3] have shown that external factors, such as ambient temperature, can have a significant impact on the energy cost of wireless transmissions.

II. PROBLEMATIC

Wireless communication can be affected by weather changes such as rain, snowfall, wind, and particularly temperature and humidity. While temperature and humidity have insignificant effect on radio wave propagations on the air, they significantly affect radio equipments [5] , [6] . By affecting the RF (Radio Frequency) circuitry, high temperatures leads to the deterioration of the radio link quality.

It is well established that temperature affect radio both transceivers at the receiver and the transmitter side. Information given in the data sheets of widely used low power radios (such as the CC2400) shows that both the sensitivity of and the output power decreases with increasing temperature [4] see figure 2 .

There is also increasing evidence from within the research community that high temperature affect the transmitted signal at the transmitter side. According to [5], high temperatures affect the transmitter by causing distortion in the output waveform during final stage amplification, which results in a poor signal being transmitted. In addition, the authors of [6] showed that there is a decrease in link quality that can reach 8dB over a temperature range of 40 C.

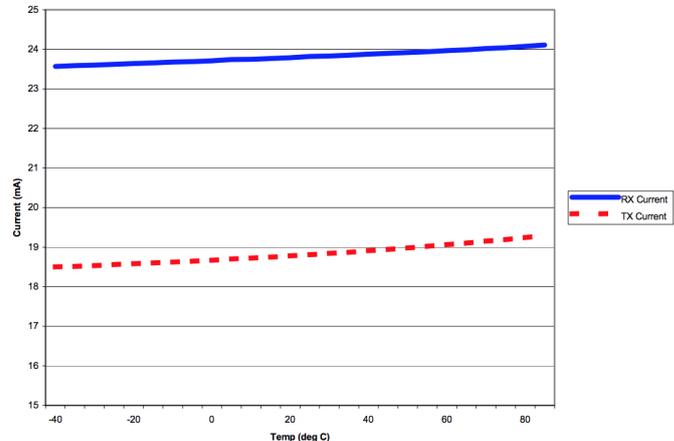


Fig. 1. Typical RX and TX current vs. temperature [1]

A. Effect of Temperature on Overall Energy Consumption

1) Network model: Consider a network of n sensor nodes $1; 2; \dots; n$ and a base station node labeled $n+1$ distributed over a region. The locations of the sensors and the base station are fixed and known a priori. Each sensor produces some data as it monitors its vicinity. We assume that each sensor generates one data packet per time unit to be transmitted to the base station. For simplicity, we refer to each time unit as a round. We assume that all data packets have size k bits. The information from all the sensors needs to be gathered at each round and sent to the base station for processing. We assume that each sensor has the ability to transmit its packet to any other sensor in its radio range. Further, each sensor i has a battery with finite, non-replenishable energy E_i . Whenever a sensor transmits or receives a data packet, it consumes some energy from its battery. The base station has an unlimited amount of energy available to it.

2) Drained power per bit : Each node drains some energy to transmit and receive data. In order to evaluate this energy with respect to node's activity, we approximate the drained energy per bit in transmit ($E_{TX}(T_t)$) and receive ($E_{RX}(T_r)$) modes as follows :

$$E_{TX}(T_t) = \frac{P_{TX}(T_t)}{LS} \text{ (Joule/bit)} \quad (1)$$

$$E_{RX}(T_r) = \frac{P_{RX}}{LS} \text{ (Joule/bit)} \quad (2)$$

where LS is the link speed . Contrary to the famous energy model proposed by Heinzelman et al. in [7], we consider in

this model the temperature impact on the drained energy per transmitted or received bit.

3) Drained power per node: The drained power by a node depends on its activity: sensing, transmitting, receiving, sleeping, listening, etc. For the sake of simplicity, we consider only the drained power due to transmitting and receiving. Let $f_{i,j}$ be the total number of bits that node i transmits to node j . It follows that for each node $i = 1; 2; \dots; n$, the total drained power can be written as follows:

$$\sum_{j=1}^{n+1} f_{i,j} \times E_{TX}(T_t) + \sum_{j=1}^{n+1} f_{j,i} \times E_{RX}(T_r) \quad (3)$$

Note that T_t and T_r are functions of time.

4) System lifetime and optimization problem: We rely on the definition of a lifetime T of the system introduced by Kalpakis et al. in [8]. T is defined to be the number of rounds until the first sensor is drained of its energy. A data gathering schedule specifies, for each round, how the data packets from all the sensors are collected and transmitted to the base station. For brevity, a data gathering schedule is referred to as a schedule. Observe that a schedule can be thought of as a collection of T directed trees, each rooted at the base station and spanning all the sensors i.e. a schedule has one tree for each round. The lifetime of a schedule equals the lifetime of the system under that schedule. Clearly, the system lifetime is intrinsically connected to the data gathering schedule.

The optimization problem can be written as follows: The first constraint is to respect the available energy at each node. The second constraint is for flow conservation. The third one guarantees to not exceed the maximal flow on each link, and the fourth constraint guarantees that the generated flow at each node reaches the base station.

III. NUMERICAL RESULTS

We would like to evaluate the performance of our models in a fix square grid of 200x200 points. We assume that the sensor nodes are randomly deployed on area and the sink can be located randomly inside the area.

We compare numerically results for various networks with different sizes for our model with Dijkstra Shortest Path Model Fig 2 represent the averages of the results obtained from 10 randomly generated network instances for each network size considered. We can observe that our model have more life time compare to Dijkstra model and that as the network size increases the lifetime of model also increases.

IV. CONCLUSION

My current activity focuses on finding a schedule that maximizes the system lifetime T taking into consideration the impact of temperature in both sending and receiving data in function of the distance that separates the transmitter from the receiver to help us find the optimal deployment that achieves the minimum overall mean power consumption.

Concurrently, I am studying the temperature impact on radio link quality to determining the best transmission power

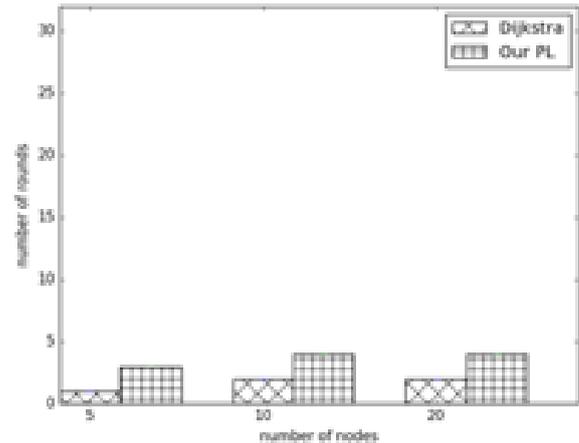


Fig. 2. numerical results

to use in order to maintain network connectivity and metrics to get a better estimate about the relation between temperature and link quality.

At the same time, I am studying the simulation in Python and omnetpp .

The main objective of this work is to identify all possible levels to optimize energy consumption and develop solutions to maximize the life of IoT communications: a communication to be robust in the broad sense, reliable and efficient while tolerating possible faults due to the depletion of energy resources.

REFERENCES

- [1] International Electrotechnical Commission "Internet of Things: Wireless Sensor Networks" White Paper 2014.
- [2] Sandra Sendra, Jaime Lloret, Miguel García and José F. Toledo, "Power saving and energy optimization techniques for Wireless Sensor Networks", Journal of Communications 6(6):439-459 ,September 2011.
- [3] Abdelmalik Bachir, Walid Bechkit, Yacine Challal, Abdelmadjid Bouabdallah, "Joint connectivity-coverage Temperature-aware algorithms for Wireless Sensor Networks", IEEE Transaction on Parallel and Distributed Systems, pages: 1923-1936, 2015.
- [4] C. Corporation, "CC2400 2.4 GHz Low-Power RF Transceiver, Data Sheet," 2008. Available: <http://www.ti.com/lit/gpn/cc2400>.
- [5] J. Otero, P. Yalamanchili, and H.-W. Braun, "High performance wireless networking and weather," University of California, San Diego, Tech. Rep., 2001.
- [6] K. Bannister, G. Giorgetti, and S. Gupta, "Wireless sensor networking for hot applications: Effects of temperature on signal strength, data collection and localization," in Proceedings of the 5th IEEE workshop on Embedded Networked Sensors, Charlottesville, VA, USA, 2008.
- [7] W. R. Heinzelman, A. Chandrakasan, , and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000, pp. 1? 10.
- [8] P. N. Konstantinos Kalpakis, Koustuv Dasgupta, "Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks," Computer Networks, vol. 42, p. 697716, 2003.

Considering learners' emotions in Serious Games for eLearning context: a new approach

MOSTEFAI Belkacem

Directeurs de thèse : Pr BALLA Amar (ESI) & TRIGANO Philippe (UTC/France)

Équipe : EIAH

Résumé— Serious games (SGs) are games explicitly designed with educational purposes. In eLearning context, the study of learner's emotions within SGs games is still relatively novel. During the game playing session, learners' emotions are continually changed by the game events, but serious games do not adapt to learner's emotions. In this paper, we present a new efficient approach for considering learner's emotions in serious games adaptation. In order to evaluate our approach, we designed and implemented a SG called DAPM, and then we conducted an experiment with learners by using tow version of DAPM: a static game version and an adapted game version. Significant differences were found between learners however pointing to the importance of taking into account the learner's emotions in SG adaptation.

I. INTRODUCTION

The focus of current research in SGs area is the consideration of player's emotion in game adaptation. The main goals of our approach is to propose a generic modelling technique for developing serious games, which must be able to estimate in real time the emotional states of a learner and adapt its reaction according to these emotions. The process of adaptation must be take into account the positive and négative aspect of emotion. Positive emotions have a positive influence on learner-player's behavior and engagement, while negative emotions have a negative influence impact on them) "in press" [1].

In this paper, we will first explain our approach. Then we will present DAPM scenarios and its adaptation module. Finally, procedure for experiment with DAPM and approach's evaluation are described.

II. A NEW MODELING APPROCH

The aim of our modeling approach is to determine for any SG, a method for estimating, in real time and outside the physical constraints, the learner's emotional states during playing SG, and for adapting, in real time, the SG's interactions according the changing of player's learner's emotional states. In our approach, we define SG as input/output dynamic system. Inputs are all the possible actions of the player, while outputs means the game's events provided by the game mechanics. This vision is inspired form a definition of educational game given by ZILAS [2].

A. Emotion recognition in SG

Several researcher shows that emotion is a main factor that give a player the tendency to perform particular action over others. In [3], authors suggest that many emotions can be defined by different modes of action tendency. Perron [4] use link between emotion and action tendency to characterize some prototypical gameplay emotions. actions) which is guided by both emotions and goals. Mick Power and Tim Dalglish [5] suggest an emotion model where different basic emotions are linked to different goal status evaluations and the their action tendencies. From all these precedent finding, we

construted an algorithm in order to estimate, in real time, the emotional state for a palyer, when he/she playnig a specific SEG.

Algorithme : real-time estimation of emotional state in in SEG

```

01 : Declaring and Initializing Variables and Data Structures :
    Init (Current_emotion, Current_action, Current_goal_state)
    Events = {v1, ..., vi, ..., vn}      Emotions = {e1, ..., ei, ..., em}
    Actions = {a1, ..., ai, ..., am}     goalstatus = {s1, ..., si, ..., sp}
    Action_tendencies = {(ei, ai, ..., aj) | ei ∈ Emotions, aj ∈ Actions}
    E_SA = {(ei, sj, ak)}, ei ∈ Emotions, sj ∈ goalstatus, ak ∈ actions
02 : Daclarng : P(aik) = {e1 ... e1}, aik ∈ Actions, ei ... ej ∈ Emotions
03: repeat
04: If an event vi ∈ Events occurred then
05: S ← Current_goal_state
06: A ← Current_action
07: EAC ← F(A)
08: For each e ∈ EAC do
09: For each triplet (ei, sj, ak) ∈ E_SA
10: if (e = ei and S = sj) then
11: Current_emotion = ei;
12: End_if
13: End_For
14: End_For
15: end_if
16: until end_game
    
```

Figure 1- Estimation of emotional state based on action and goal status

B. Adaptation in SG

Player's emotional states can unconsciously change from one state to another according to one or more events. Figure 1 describe symbolically the change of player's emotional state, at a specific step in time during the playing process. Thus, we use markov chain approach to model the changes in player emotional state for each event occurred when playing game.

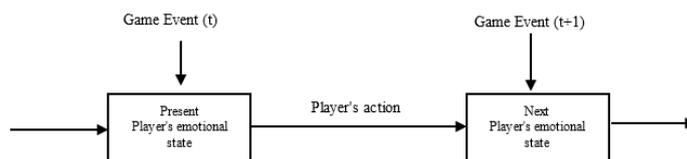


Figure 2 the change of player's emotional state according to game events

The modeling of player's emotional states as a finite-State Markov Model allows us, through empirical tests, the consruction of a transition matrice for each events in the context of a specific serious game . The transition matrix with N emotional states is an $N \times N$ matrix where the entry in row i and column j is the probability that a player's emotional state in emotion S_i will be turn on emotion S_j on the next event. Thus, based on a transition matrix, we propose an algorithme for adapting the game behavior according the the changes of emotionals player states.

The algorithm of emotional adaptation will be performed at each significant time interval. So, if the current goal not achieved, this algorithm verify if the number of positif emotions is less then the number of negatif emotion. If so, it identify the emotional state corresponds to this action by calling a function which allow estimation of emotion from action and the state of the current goal. Then it adapt the next events for inducing positif emotions or

reducing negative emotions. Otherwise, the adaptation is not necessary.

```

Algorithm : adaptation in SG based on learner's emotions
01: repeat
02:  ne = number_of_negative_emotion();
03:  pe = number_of_positive_emotion();
04:  if ( Current_goal_state not achieved and ne>=np)
05:    enow = Estimation_Current_Emotional_State();
06:  repeat
07:    select vk a next event from next_possible_events().
08:    select enext from setof_positive_emotion();
09:    select PEvk/personalityType&player_type = [Pi]
10:  until Evk(enow|enext,vk) > 0
11:  event_adaptation( enow,vk, enext);
11: until end_game.
    
```

Figure 3- game game

This algorithm use both personality type and playe type as a basis for improve our algorithme of adaptation. For each pair of personality and player type, we construct gtransition matrix related each event of the SG.

III. DAPM

In order to evaluate our approach, we designed and developed a SG named DAPM (Drone aginast Programmable Machines). DAPM is implemented in two versions, one is static then the other is an adapted game.

A. DAPM : goals and scenarios

DAPM is a training game where the player control a military drone to eliminate programmable machines and protect manufacturers from destruction. DAPM missions are concerning and challenging player's knowledge about the training of simple programs written in a specific programming language (training goal) and requiring resource management ability in the use of drone weapons resources (fun goal). DAPM scenarios are about achieving mission objectives for each of the six game levels. To achieve the objectives for a specific level, the player must be able at the same time to maneuver the drone to distract the enemy on the one hand, and from the other and he/she use the drone's radar for display a tank program and analyze it to obtain important information (intensity of its resistance,...), and then the use of weapons drone intelligently to destroy tanks.

B. Assessment of Emotions in DAPM

At first, we develop an non adaptive version fo our game DAPM. Afterwards, we conceive tow algorithms, the firste one is for the recognition (estimate and tracking in real time) of player's emotional states, the second one (see section) is for the adaptation of game behavior based on events-emotions pattrens.Theses algorithms are embedded within game adaptation module, their implementation is heavily dependent on the game under examination. Thus, we present below the five steps followed in order to implement theses tow algorithmes into DAPM game :

Step 1- by using self-reports method, we have identified a set of different positive and negative emotional states experienced by different players when playing the static game version of DAPM. we used AEQ-M questionnaire [6] and we constrcut the second questionnaire based on LAZZARRO work [7], Here is the list of player's emotions during DAPM game play : interest , joie, engagement, fiero, relief , surprise ,boredom, frustration , fear, anger, pride, anxiety.

Step 2 - the obtained emotional states in step 1 are linked to diffrents player's action tendencies and different goal status in

order to use it in the implementation of our player's emotions recognition algorithm. Table show a part of the reationship between this player actions, goal status and emotional states.

Step 3 – We have implemented two questionnaires which the player must answer when creating his account game. We use Lazzaro's players type questionnaire for classifying players according their behaviors. We adopt famous Eysenck Personality Questionnaire Revised-Abbreviated (EPQR-A) on order to calssify player's personality trait.

Step 4 : we conductes une exprimntation with 40 students for the construction of transition matrix for each pair of player personality and player type. Each matrix contains transition probabilities values of emotional states according each game event. The data structure of transition matrix are used in the implementation of our algorithm for game adaptation.

Step 5 - in this step we implemen the adaptaion function that is responsible for determining how and when adapt game events for increasing positif emotion at player. this funtion exploit the transition matrix for a smart management of next events when A current goal not achieved and the number of player's negative emotional states is greater than positive ones.

IV. EXPERIMENTATION, RESULTS AND EVALUATION

We conducted an experiment with two groups of players. Each group is composed of 25 undergraduate students. The first group is asked to play non-adaptive version of the game, then the second group one is request to play the adaptive one. All players are invited to play and try to complete the game for up to six days. Then each one of them provided us with a detailed feedback about his/her user experience. Both versions of games are implemented as web game. For each game version they are a database in which we have saved values for some indicators, such as numbers of positive / negative emotions, the number of failed attempts before completing a given level. The results obtained by questionnaires and by analyses of indicators demonstrate that the adapted game version creates more fun at the players than the non-adaptive game. Furthermore, adapted game version brings better performance of the majority of players.

V. PERSPECTIVES

Based on our generic approach, we can developing and including an emotional adaptation module in any SG type. We have made an experimentation to valid our approach. The result of the experiment shows the player's emotions influence the playing and the learning activities in the context of the game. The adaptation module seems very interesting as it enables the game to acclimate its events for avoid negative emotion or to induce positive one.

References

- [1] B. MOSTEFAL, A. Balla, and P. Trigano , “positive and negative emotions of learners in serious games: an analytical study ”, CELDA 2017 “paper in press”, website : <http://celda-conf.org/>
- [2] ZILAS, Nicolas, ACOSTA, Martin Eduardo. A Theoretical Background for Educational Video Games. In: P. Felicia. Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. Hershey, PA : IGI Global, 2011. p. 215-238.
- [3] N.h Frijda. the emotions. cambridge university press. 1986.
- [4] Perron, B. 2005. A cognitive psychological approach to gameplay emotions. Conference Proceedings of DiGRA.
- [5] D.R. Michael and S.L. Chen, “Serious Games: Games That Edu- cate, Train, and Inform”. Muska & Lipman/Premier-Trade, 2006.
- [6] Pekrun, R., et al, 2011. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). Contemporary educational psychology, 36(1),pp 36-48
- [7] Nicole Lazzaro. (2004, March, 8). Why We Play Games: Four Keys to More Emotion Without Story.

Etude de mouvement dans une séquence vidéo

Prénom Nom BENALIA Mohcene

Directeurs de thèse : Professeur Samy. Ait-Aoudia

Équipe : IMAGE - Laboratoire de recherche : Méthodes de Conception de Systèmes (LMCS)

Résumé— Dans ces dernières années, le domaine de vision par ordinateur a connu un intérêt par plusieurs applications ; par exemple les applications visuelles sur Smartphones, surveillance, compression ou indexation vidéo. Toutes ces applications utilisent des méthodes où la première étape c'est la détection de mouvement. Cependant, l'implémentation des méthodes de précision, qui donnent des bons résultats en temps réel, reste un challenge selon les nouveaux besoins des ces applications. Dans notre étude, nous avons proposé une nouvelle approche qui est inspirée de la modélisation par mélange des gaussiens (Gaussian Mixture Model). L'objectif principal de la méthode proposée est d'accélérer le temps de traitement et de garder la qualité des résultats ou des les améliorer. Afin de réaliser ces objectifs, nous avons réduit le flux reçu des images par des échantillons. Par la suite, une estimation multi gaussiennes est faite par les pixels reçus. Alors, le fond dynamique sera construit par les distributions les plus fortes et le mouvement par celles les plus faibles. Des itérations par le Framework de MRF (Markov Random Field) sont appliquées sur les résultats pour les améliorer.

I. INTRODUCTION

La vision par ordinateur ou la vision artificielle vise à simuler la vision naturelle avec ses différentes fonctions en performance qui doit dépasser celle de la vision naturelle. Ce domaine a connu un intérêt remarquable, ces dernières années, qui revient au développement technologique nécessitant des applications visuelles; par exemple : Surveillance et suivi des objets en mouvement

Ce domaine comporte plusieurs axes de recherche : amélioration, reconnaissance, compression, indexation vidéo, etc. Entre ces axes, la détection de mouvement dans une séquence vidéo représente un prétraitement (étape primaire) pour plusieurs sujets par exemple : la compression vidéo, la surveillance et le contrôle de trafic routier.

Dans cet axe de recherche, plusieurs méthodes ont été proposées afin d'améliorer la détection du mouvement sous la présence des plusieurs challenges. Généralement, on peut classer ces méthodes selon la nature du fond en deux classes :

- Fond statique : le fond ne change pas dans un intervalle de temps, alors chaque position du fond est représentée par une seule information (un pixel). Les méthodes de cette classe sont orientées pour des applications à l'intérieur indoor.
- Fond dynamique : le fond se change dans un intervalle de temps, alors chaque position de ce fond est représentée par plusieurs informations (plusieurs pixels). Les méthodes de cette classe sont orientées pour des applications à l'extérieur

Autre que la qualité des résultats, les méthodes de détection de mouvement cherchent l'adaptation aux certaines circonstances ; par exemple :

- Élimination des ombres
- Change graduel de luminosité
- Change soudain de luminosité
- Un objet stationnaire en réveil
- Un objet mobile en stationnement

Dans notre projet, nous avons proposé une nouvelle méthode, de détection de mouvement et de construction de fond, qui représente le cœur de notre sujet. De plus, nous avons proposé des améliorations et des méthodes dans le même axe qui seront détaillée dans la partie des travaux réalisés

II. PROBLÉMATIQUE

La problématique principale qui se pose, dans la détection de mouvement, c'est la bonne qualité du masque de l'objet en mouvement sur le critère de temps du calcul pour un fond dynamique. Le gain en ces trois objectifs reste un challenge à combattre, qui reste demandé par une certaines applications en temps réel. Notre problématique de recherche posée est d'offrir une méthode (modèle ou des outils) utile à utiliser pour les niveaux supérieurs en vision par ordinateur. La méthode doit être flexible et non liée à une application donnée. Autrement, notre axe de recherche ne représente pas l'application finale en vision par ordinateur, mais il peut être vu comme une pré-étape essentielle pour les applications en surveillance

III. TRAVAUX REALISES

(les mémoires encadrés et les articles ou les communication sont cités à la fin de cette section)

Les travaux réalisés peuvent être cités dans les points suivants :

1. **Proposition d'une nouvelle méthode de détection de mouvement et de construction du fond** : qui représente l'objectif de notre sujet. La méthode est proposée avec deux implémentations différentes qui se basent sur l'aspect de la modélisation par mixture des gaussiennes. Les résultats ont prouvé les performances en qualité du masque de mouvement pour l'implémentation SGM et en temps de calcul en ce qui concerne l'implémentation QSGM.

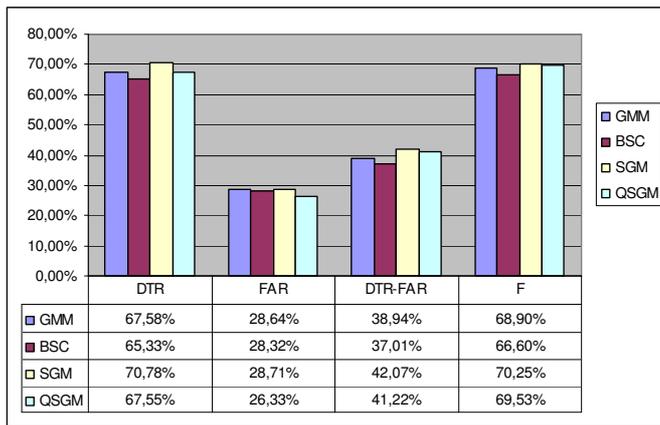


Fig.1 : résultats de détection de la méthode proposée (SGM, QSGM) comparés avec les résultats des autres méthodes (GMM, BSC)

2. **Détection des mouvements lents dans une séquence vidéo:** Nous avons utilisé la modélisation markovienne (champ de Markov aléatoire MRF) pour la détection des mouvements lents qui ne sont pas possible de les détecter avec les méthodes classiques

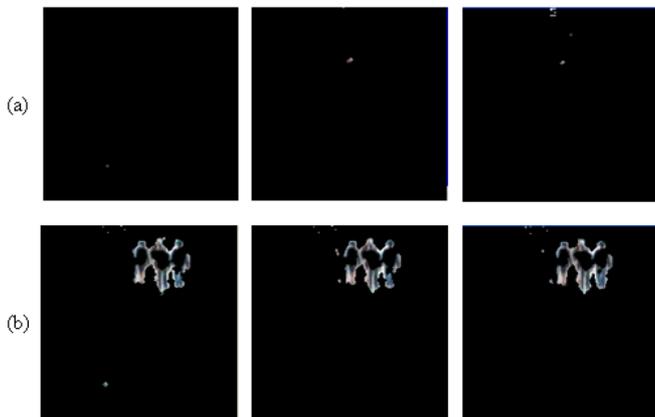


Fig.2 résultats de la détection avec la méthode proposée (b) avec les résultats de la modélisation classique (a)

- Amélioration de la méthode BSC :** le travail a été publié dans une communication [1]
- Détection de mouvement dans des vidéos en couleurs :** Afin d'améliorer la qualité de l'objet détecté en mouvement, nous avons utilisé les trois canaux des couleurs sur les méthodes basant décision pixel.
- Implémentation d'un système de surveillance :** nous avons amélioré le temps d'alerte, du système de surveillance pour la détection des objets stationnaires, par l'utilisation de la méthode de détection BSC.

A. *Communications et articles publiés ou en cours de rédaction*

- "An Improved Basic Sequential Clustering Algorithm for Background Construction and Motion detection"; ICIAR 2012, LNCS. http://rd.springer.com/chapter/10.1007/978-3-642-31295-3_26 cité 7 fois [1]

- "Markov random field for slow motion detection and object segmentation"; ISPA 2017, Mostaganem Algeria en soumission (<http://ispaconf.univ-mosta.dz/index.php>)
- "New multi Gaussian sampling form motion detection and background construction" , un article en soumission. [2]
- "An improved BSC and GMM for motion detection by color decision" un article en rédaction
- "Basic sequential clustering for abounded object detection", un article en rédaction. [3]

B. *Contributions*

- Nous somme cités dans la partie - acknowledgement- de l'article de M Nawaz «Precise foreground detection algorithm using motion estimation, minima and maxima inside the foreground object »
- Nous somme cités dans la partie de remerciement du mémoire PHD de M Nawaz « Foreground detection of video through the integration of novel multiple detection algorithms »

C. *Mémoires encadrés dans l'axe de recherche:*

"Manoeuvred visualization of 3D objects in augmented reality" 2017

IV. PERSPECTIVES

Dans ce qui va suivre, nous allons soumettre les travaux rédigés dans des revues et dans des conférences. Comme nous allons compléter l'étude de la parallélisations des méthodes de détection basées pixel (exemple GMM BSC) et d'améliorer les performances en temps d'exécution de l'implémentation des méthodes pour les vidéos en couleurs

V. CONCLUSION

Dans notre sujet, nous avons proposé une nouvelle méthode qui est inespérée de la modélisation Gaussienne. Les résultats obtenus prouvent la puissance de notre approche, où elle a donné des résultats améliorés, en qualité de l'objet détecté et en temps de calcul, en comparaison avec d'autres méthodes. La méthode peut être améliorée en éliminant les ombres. Dans un autre part de l'axe de recherche, l'outil de parallélisme va devenir un point d'intérêt pour les méthodes basant pixel qui revient à la nature simple des données de ces méthodes.

References

[1] M. Benalia, S Ait-audia "An Improved Basic Sequential Clustering Algorithm for Background Construction and Motion detection" ICIAR 2012, LNCS

[2] M. Benalia, S Ait-audia "New multi Gaussian sampling form motion detection and background construction".

[3] M. Benalia, S Ait-audia "Basic sequential clustering for abounded object detection".

Communications security in the next mobile networks generation

NAIT HAMOUD Othmane

Directeurs de thèse : KENZA Tayeb, CHALLAL Yacine

Équipe : SURES

Abstract—Device-to-Device (D2D) communications is a promising technology for the next generation mobile communication networks (5G). It is expected to allow high throughput, reduce communication delays, energy consumption and traffic load and thus enhance traditional cellular networks' capacity and performance. The standardization of this technology is underway by the Third Generation Partnership Project (3GPP) under the proposal Proximity Services (ProSe) [7] which allows enabling direct communication between proximate devices without passing through an access point or a core network of a cellular infrastructure. However, due to the open nature of wireless communications, D2D communications face several threats. Security issues must be considered in this technology, especially when it comes to wireless communication between devices involved in controlling critical infrastructures and/or dealing with private data. We are interesting in the thesis to contribute in D2D communication security.

I. INTRODUCTION

The rapid growth in the number of mobile internet subscribers has fostered the emergence of various new applications and services. This implies an exponential growth of mobile data traffic. Consequently, a huge burden is imposed for the cellular infrastructure in terms of spectrum utilization, overall throughput, communications delays and energy consumption.

Expected to be one of the technology components of the evolving 5G architecture, Device-to-Device (D2D) communications is a promising solution to offload the cellular infrastructure from the traffic encumbrance. Indeed, D2D communications approach allows devices (device such Smartphone, tablet, etc.) to establish direct communication links with each other without passing through an access point or a core network of a cellular infrastructure. The main difference between the expected 5G and the first four generations is that 5G is heading towards device-centric network architecture contrary to the previous generations which have been network centric. In 5G, a device is expected to actively perform operations which were earlier being performed by the network such as storage, relaying and content delivery [1].

These recent years, academic, industrial, and standard institutions have paid considerable attention for D2D communications technology. In academia, different surveys have been proposed in the literature [1], [2], [3], [4], [5] in which, different fields related to this technology was addressed (node discovery, interference and radio resource management, use cases and requirements, power control, system architecture and design, etc.).

In industry, Qualcomm has developed FlashLinQ [6] to implement for the first time D2D communication as sub-system underlying cellular networks to enable direct com-

munications among proximity devices in different scenarios (content sharing, gaming, social networking, etc.). FlashLinQ was designed to work in licensed cellular band based on Time Division Duplexing-Orthogonal Frequency Division Multiple Access technology (TDD-OFDMA) which is the same as LTE-A system, allowing devices to discover neighbors in a large range with high efficiency.

The standardization of this new paradigm is underway by the Third Generation Partnership Project (3GPP) under the proposal Proximity Services (ProSe) [7] which allows enabling direct communication between proximate devices. ProSe combines two types of services, proximity discovery and direct communication. In [8], a brief overview of standardization activities of the 3GPP ProSe in LTE-A is presented.

II. PROBLEMATIC

D2D communications face many security challenges as part of the future 5G systems. The importance and the dimension of these challenges are to define depending on many factors: open air nature of wireless communications, large-scale applications, use cases and scenarios, adoption of D2D technology by users at a large scale, pricing and business models, etc.

Security in D2D communication is still in an embryonic state. Few works have handled security issues in this novel technology. Furthermore, these works address security in a scattered way as each solution is defined in a well-defined scenario and faces well-defined threats and does not attempt to solve security problems in their entirety. Generally, existing security solutions work on a specific layer (application, network, MAC or physical layer), while the few solutions which consider corresponding security aspects of more than one layer do so only independently. We are interesting in this thesis to develop a security solution which fulfills all security requirements, faces all security threats and supports all D2D communication scenarios. The approach we advocate is based on a joint framework which involves each layer security (physical, MAC, network and application) technology to work in a cooperative way to overcome efficiently security issues.

III. WORK COMPLETED

Until this stage of the advancement of this thesis, we have studied D2D security architecture, threats and requirements and proposed a taxonomy based on the review of recent works which have addressed the security issues in D2D communications [9]. The proposed taxonomy is more practical since it gives, on one hand, a better readability and a good understanding of all the works that have addressed the security issues in the literature, and on the other hand, a roadmap

towards a global security solution that combines the best techniques and security solutions inherent to each layer.

The majority of works within the application layer deals with cryptographic key management issues in order to apply them in a specific context. From the cryptographic point of view, key management schemes are important to find efficient cryptographic solutions in order to satisfy requirements in terms of authentication, confidentiality, integrity and many other security issues. The proposed solutions in the literature didn't assume all scenarios related to the involvement of cellular infrastructure (i.e. assisted, controlled or autonomous), the most important difficulties concern keys distribution and revocation problems.

D2D communications may gain advantage from the control and the assistance of a cellular infrastructure by getting necessary credentials to be employed in case of intermittent cellular connectivity or out of coverage scenarios. The operator's core network can ensure the role of certification authority in Public Key Infrastructure (PKI). However and contrary to the goal of introducing D2D technology in cellular networks (in offloading scenario), this one is burdened by certificates management. Furthermore, in the out-of-coverage scenario, D2D communications could occur without a trust authority. Identity-Based Cryptography (IBC) was introduced to overcome and simplify the public keys management [10]. However, IBC suffers from the key escrow problem since the trust authority generates user's private key and consequently, can impersonate user's identity and decrypts its encrypted data. Certificate-less Public Key Cryptography (CL-PKC), which is intermediate between traditional PKI and IBE, was conceived to overcome the aforementioned two problems (i.e. certificate management and escrow problem)[11]. we aim to introduce CL-PKC in such away that it extract secret keys from the measurement of physical layer and provide cryptographic credentials to secure multi-hop in D2D communications.

IV. PERSPECTIVES

Generally, D2D communications rely on one hop routing; however in different scenarios (public safety, extension of coverage, dissemination of content, etc.) they may rely on multi-hop routing. Few works in the literature have treated routing aspect in D2D communications. From the security point of view, much work remains to be done, especially to face security threats related to the absence of trust authority, and the highly dynamic of the topology on one hand; and on another hand to preserve security and privacy of users which will see their sensitive information transit different nodes without trust authority. Besides, malicious contents can be injected into the D2D network and affect UEs with viruses, malwares and many other threads. Secure D2D routing through a cryptography approach needs manipulating cryptographic keys that key management schemes must take into account. Another approach to secure routing in D2D communications relies on network coding which employs data splitting and shuffling mechanisms over butterfly networks.

Otherwise, physical layer security is playing a key role for securing wireless communications in recent years. It exploits physical characteristics of wireless channel to prevent essentially from eavesdropping attack without utilizing cryptographic approaches. Works related to this field turn around

theoretic secrecy capacity, CSI-based authentication and CSI-based key agreement.

We plan in the future to design a cross-layer security architecture to overcome efficiently security issues in D2D communications.

V. CONCLUSION

D2D is a promising technology in LTE-A networks. Taking advantage of proximity devices, it offers high throughput, lower delays and offloading cellular networks traffic. On the other side, it offers a variety of practical services (advertising and commercial services, public safety services, etc.).

There are many design challenges in D2D so that much research effort is still needed. By reviewing many of recent works related to security in D2D communications, we notice that these works are scattered depending on some specific security issues in different security aspects and contexts.

Through this work, we have shed-light on the necessity to develop a global solution that combines the best techniques and security solutions inherent to each layer.

REFERENCES

- [1] P. Gandotra, R. K. Jah, and S. Jain, *A survey on device-to-device (D2D) communication: architecture and security issues*, Journal of Network and Computer Applications 78 (2017), pp. 9-29.
- [2] A. Asadi, Q. Wang, and V. Mancuso, *A survey on Device-to-Device communication in cellular networks*, IEEE Communications Surveys and Tutorials, vol. 16, no. 4, pp. 1801-1819, April 2014.
- [3] J. Liu, N. Kato, J. Ma and N. Kadowaki, *Device-to-Device communication in LTE-Advanced networks: A survey*, IEEE Communications Surveys and Tutorials, vol. 17, no. 4, pp. 1923-1940, December 2015.
- [4] Pimmy Gandotra and Rakesh Kumar Jha, *Device-to-device communication in cellular networks: A survey*, Journal of Network and Computer Applications, <http://dx.doi.org/10.1016/j.jnca.2016.06.004>
- [5] G. Fodor et al., *Design Aspects of Network Assisted Device-to-Device Communications*, IEEE Commun. Mag., vol. 50, no. 3, Mar. 2012, pp. 170-77.
- [6] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic, *FlashLinQ: A synchronous distributed scheduler for peer-to-peer ad hoc networks*, IEEE/ACM Transactions on Networking, vol. 21, no. 4, pp. 1215-1228, June 2013.
- [7] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility study for Proximity Services (ProSe) (Rel 12), 3GPP TR 22.803 V1 2.2.0 (2013-06).
- [8] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, *An overview of 3GPP device-to-device proximity services*, IEEE Communications Magazine, December 2013.
- [9] O. Nait hamoud and T. Kenaza and Y. Challal, *Security in Device-to-Device (D2D) Communications: A Survey*, under consideration for publication in IET Networks Journal, 2017.
- [10] R. H. Hsu and J. Lee and T. Q. S. Quek, and J. C. Chen *GRAAD: Group Anonymous and Accountable D2D Communication in Mobile Networks*, CoRR abs/1703.04262 (2017).
- [11] Al-Riyami S, Paterson K, *Certificateless public key cryptography*, Advances in cryptology-ASIACRYPT. Lecture notes in computer science, vol 2894. Springer, New York, pp 452-473, 2003.

Analyse d'images satellites et photos aériennes multicritères pour la détection de changement sur des scènes urbaines : application à la mise à jour des bases de données géographiques .

Hayet Si Salah

Pr Ait Aoudia Samy. Pr Bachari Nour El Islam

Résumé— Les images satellites et les photos aériennes représentent des sources de données hétérogènes de plus en plus nécessaires et difficiles à exploiter. Ces images sont considérées comme des sources très riches et utiles pour le suivi de l'environnement par détection de changement.

L'objectif de cette thèse est de développer une nouvelle méthode de détection de changement par analyse multi critères de photos aériennes et images satellites.

I. INTRODUCTION

De nos jours, les images satellites et des photos aériennes représentent des sources de données hétérogènes de plus en plus nécessaires et difficiles à exploiter. Ces images sont considérées comme des sources très riches et utiles pour la mise à jour des bases de données géographiques, cette dernière nécessite une étape de détection de changements qu'il faut développer à partir de ces images et photos multi critères (multi temporelles, multi résolutions, multi capteurs, ...etc).

II. PROBLÉMATIQUE

La mise à jour des bases de données géographiques concerne plusieurs couches d'informations avec des fréquences de changement variées, la couche planimétrie qui représente l'urbain subi généralement beaucoup de changement vu les effets de la nature et des projets de l'état. Elle nécessite donc une mise à jour périodique.

L'objectif de cette thèse est de développer une nouvelle méthode de détection de changement par analyse multi critères de photos aériennes et images satellites à haute résolution dans le but de mettre à jour la couche planimétrie de la base de données géographique. Cette méthode repose sur un modèle de classification suivi d'un modèle d'appariement.

Une étude bibliographique élargie doit être faite à fin de pouvoir arriver à cet objectif et choisir les procédures techniques adéquates.

Le travail que nous devons mener dans le cadre de cette thèse de Doctorat est prévu selon l'échéancier global suivant :

- Faire une recherche bibliographique et un état de l'art du domaine considéré (1 an).
- Étude critique des différentes méthodes proposées .Développement d'une méthode pour la problématique étudiée (1 an).
- Exploitation des résultats obtenus (comparaisons et conclusions) (1 an).

- Publication dans une revue et rédaction de la thèse (1 an).

III. TRAVAUX RÉALISÉS

Depuis la première inscription faite en mois d'octobre 2014 et acceptée par le CS en janvier 2015 (notre point de départ), nous avons pu terminer l'état de l'art (recherche bibliographique) ou nous avons étudié quelques thèses, articles du domaine considéré, nous avons fait une synthèse (points forts et points faibles) pour chaque travail. Une révision et mise à jour récente de l'état de l'art a été également faite durant la période de stage octroyé vu l'accès facile a une documentation plus riche. Un article [1], a été rédigé (en cours de publication dans un journal).

Nous avons également fait la récolte des données nécessaires pour l'évaluation de notre future travail (images satellites, données vectorielles,... Etc.).

Une contribution concernant la proposition d'une nouvelle méthodologie pour la détection de changement a été initiée .Dans ce cadre, un deuxième article [2] est en cours de finalisation.

Un troisième article comprenant l'architecture et l'implémentation de notre nouvelle solution proposée est en cours de réalisation.

IV. PERSPECTIVES

Soumission et publication de l'article [1] finalisé.

Finaliser l'article 2 dans le plus bref délai et le soumettre pour publication.

Terminer la partie expérimentation durant un stage (qui débutera à la fin du mois de Septembre 2017) dans une société de domaine pour évaluer les résultats de la méthode dans un environnement de travail réel. Permettant ainsi de finaliser l'article 3 et le soumettre pour publication.

Poursuivre la rédaction de la thèse.

Aussi la participation dans des conférences et événements scientifiques si ça sera possible.

V. CONCLUSION

Nous comptons finaliser tous les travaux de thèse en quelques mois tout en respectant l'échéancier établi.

References

“unpublished” [1] “Change detection in urban area from remote sensing data -review article.”

“unpublished” [2] “Towards KDD framework for change detection in urban area using remote sensing data.”

“unpublished” [3] “KDD-change detection framework modeling and implementation for GIS update using remote sensing data.”

Security in Big Data

Case Study: Privacy Protection

Zakaria Gheid

Directeurs de thèse: Prof. Yacine Challal

Équipe: SURES

Abstract—Big Data systems are gathering more and more information to discover new values and patterns through data analytics. However, analyzing Big Data stores has raised security and privacy concerns, surrounding the disclosure of private sensitive information. Moreover, traditional security mechanisms that are built on crypto-systems become inadequate, inducing unaffordable computational overhead when applied to Big Data stores. Accordingly, this thesis aims to treat the privacy concern in Big Data analytics in a more efficient way. It will propose novel secure and cryptography-free methods to build privacy-preserving algorithms for Big Data analytics.

I. INTRODUCTION

Over the last years, world's generated data have grown in their scope and size, shifting from centralized processing to distributed environments, and leading so to the Big Data era. Big Data is a term used to designate large or complex data sets that are beyond the ability of traditional data processing and methods. The emergence of Big Data sets has created exciting opportunities to maximize the knowledge available for analysts, researchers and business people, allowing integrating Big Data analytics in decision making by uncovering hidden patterns, unknown correlations and other insights.

Moreover, Big Data analytics is performed through several advanced data mining techniques such as clustering and classification. Clustering consists of discovering patterns from a set of data objects, then affecting each object to the closest pattern. Indeed, this analysis task is now a keystone requirement for several areas of life like healthcare, social science, business and marketing. For instance, in cancer diagnosis we can take known samples of cancerous and non-cancerous data sets and apply clustering algorithms over patients' medical records (PMR) to identify cancerous data. From another side, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, based on a training set of data containing observations (or instances) whose category membership is known. Data classification may be used in a wide range of applications, such as Human Activity Recognition (HAR). HAR is an emerging research field that aims to mine pervasive data streams collected by wearable and implantable sensors to provide more understanding of human activities and behaviors. This may improve the quality of individual life in several aspects, ranging from daily assisted living to leisure applications.

Nevertheless, the effective integration of Big Data mining has given rise to privacy issues surrounding disclosing personal private data during analysis process in distributed environments. In fact, personal opinions, political interests, healthcare records, daily activities and other private data are

being shared between service providers to perform a distributed clustering or classification task.

In this thesis, we tackle the privacy issue in Big Data analytics over distributed environments. We propose efficient and secure basic primitives to perform multiparty computation in widely used data mining tasks, such as clustering and classification. In the remainder of this paper, we highlight the scope of our thesis, then, we present our past and current works.

II. PROBLEMATIC & OBJECTIVES

Facing the privacy issues caused by distributed analysis of data stores, many research works have been proposed, attempting to reach a partial privacy protection by adding some noise before sharing data [1]. However, the minimum error rate raised by the added noise is intolerable for Big Data applications needing a high accuracy level, such as healthcare. From another side, other proposed works [2] have implemented cryptographic schemes, such as the Paillier [3] and ElGamal [4] cryptosystems to securely share personal data for clustering or classification. Nevertheless, these cryptography-based security measurements coasted a significant degradation in response time, which is inadequate for Big Data sets [5], [6].

In this thesis, we propose a different approach that enhances both security and computation performance to cope with Big Data requirements. Using cryptography-free protocols, we aim to propose **efficient and privacy preserving primitives to be used for designing secure multiparty data analytic algorithms**. We make focus on widely used data mining tasks, such as clustering and classification and we target well-known algorithms in those fields.

III. PERFORMED WORKS

In this section, we present our performed works during three years of PhD studies.

A. First Year (2014-2015)

1) *State of the art draft*: During this stage, I have made the first readings in attempt to build a background of fundamental notions. These readings were about Big Data definition, Big Data security aspects and opened security axes.

2) *Problematic & thesis objectives*: The second main task that we have performed was delimiting the thesis scope to only one security aspect, which was the user privacy in data analytics. Our choice was motivated by the lack of works in this field despite its high importance in Big Data era.

3) *Conference paper*: By focusing on the privacy issue in Big Data analytics, I have started an advanced reading step, during which we have proposed (*II-CSP*) [7], a novel efficient and privacy-preserving protocol for similarity analysis, which is an important primitive in Big Data analytics.

B. Second Year (2015-2016)

1) *Engineer thesis supervision*: Build a document plagiarism detector based on our novel proposed similarity evaluation method [7] using Big Data technologies (MapReduce).

2) *Conference Paper (Core-Rank B)*: By pushing our reading toward the widely used data analytic tasks, we spotted a privacy breach in distributed Knn classification used for human activity recognition purpose, which is an important Big Data application. Therefore, we have proposed (*II-Knn*) [8], a private and efficient version of the Knn classifier, based on (*II-CSP+*) [8] that is an improved version of our previous secure similarity protocol [7].

3) *Conference paper (Core-Rank A)*: Going towards another data mining task, which is the data clustering, we have proposed a private and efficient multiparty additive scheme (Π -sum) that we used to sketch a secure version of the k-means algorithm (SK-means) [9]

C. Third Year

1) *Engineer thesis supervision*: Build a human activity recognition framework for e-healthcare using our private (Π -Knn) protocol [8].

2) *Conference paper (Core-Rank B)*: Still working on fundamental primitives that are used in analytic algorithms, we have proposed a novel Private and Efficient Set Intersection protocol [10], which is used to implement the association rule mining task. Our protocol copes with Big Data analytics as it does not involve cryptography. It incurs linear communication and server-side computation complexities; besides, it outperforms the hash-based solution used in practice.

3) *Conference paper (Core-Rank B)*: We have extended our previous work on The Private Set Intersection to propose a novel Food Adequacy Check application that aims to enhance shopping experience while preserving customer's privacy. We have submitted this work [11] to a good-ranked conference that will be held in April 2018.

4) *Journal Paper (Accepted by Elsevier JNCA)*: We have extended our paper on the private Knn protocol [8] by enhancing almost all its parts (related works, protocol design, security limitation, standard notations and experimental evaluations). Besides, we have introduced a novel medical application of our protocol that we have named SimilCare. It consists of a medical monitoring framework enhanced with a private tracking and recognition of patients' daily activities in aim to support the medical decision process. We have

submitted our manuscript in January 31st, 2017 and we have received the acceptance notification in September 14th, 2017 [12].

IV. CURRENT WORKS

A. *Journal Paper*: We are working on extending our Private Set Intersection protocol for a future journal submission.

B. *Thesis Writing*: As a final stage, I started writing my thesis and turning our research works into a single and coherent peace of work.

V. CONCLUSION

Big Data is a new phenomenon that has attracted attention of both academia and industry sectors. Because of its extraordinary features, it has conducted to new security challenges where data privacy comes in first. Nevertheless, making a short insight over Big Data research contributions may reveal a lack in efficient techniques that should preserve user privacy in Big Data analytics process. For these reasons, we have contributed in this axis, producing some privacy-preserving primitives that provides efficient privacy protection guarantees during Big Data analytics.

VI. REFERENCES

- [1] S. R. Oliveira and O. R. Zaiane, "Privacy preserving clustering by data transformation" *Journal of Information and Data Management*, vol. 1, no. 1, p. 37, 2010.
- [2] X. Yi, J. Willemson, and F. Nait-Abdesselam, "Privacy-preserving wireless medical sensor network" in *Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2013 12th IEEE International Conference on. IEEE, pp.118-125, 2013.
- [3] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes" in *Advances in cryptology EUROCRYPT'99*. Springer, pp. 223-238, 1999.
- [4] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms" in *Advances in cryptology*. Springer, pp. 10-18, 1984.
- [5] S. Patel and D. C. Jinwala, "Privacy preserving distributed k-means clustering in malicious model" 2013.
- [6] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era" *Network, IEEE*, vol. 28, no. 4, pp. 46-50, 2014.
- [7] Z. Gheid and Y. Challal, "An Efficient and Privacy-Preserving Similarity Evaluation for Big Data Analytics," *2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)*, Limassol, pp. 281-289, 2015.
- [8] Z. Gheid and Y. Challal, "Novel Efficient and Privacy-Preserving Protocols for Sensor-Based Human Activity Recognition," *2016 Intl IEEE UIC Conference*, Toulouse, pp. 301-308, 2016.
- [9] Z. Gheid and Y. Challal, "Efficient and Privacy-Preserving k-Means Clustering for Big Data Mining," *2016 IEEE Trustcom/BigDataSE/ISPA*, Tianjin, pp. 791-798, 2016.
- [10] Z. Gheid and Y. Challal. "Private and Efficient Set Intersection Protocol for Big Data Analytics" *International Conference on Algorithms and Architectures for Parallel Processing*. Springer, Cham, Helsinki, pp.149-164, 2017.
- [11] Z. Gheid and Y. Challal, "Private and Efficient Set Intersection Protocol for RFID-Based Food Adequacy check" *IEEE Wireless Communications and Networking Conference*. Barcelona, 2018. Unpublished.
- [12] Z. Gheid, Y. Challal, X. Yi and A. Derhab, "Efficient and Privacy-Aware Multi-Party Classification Protocol for Human Activity Recognition", *Journal of Network and Computer Applications, Elsevier*, 2017. In press.

Adaptation du processus KDD (Knowledge Discovery in Databases) pour une utilisation dans un environnement Big Data

LOUNES Naima

Directeur de thèse : Pr R.CHALAL

Co-Directeur de thèse : Pr W.K.HIDOUCI

Équipe : (MSI)

Résumé— Ce rapport donne d'un aperçu sur l'état d'avancement de la thèse de doctorat durant la deuxième année, sous le thème : **Adaptation du processus KDD pour une utilisation dans un environnement Big Data**. Ce rapport est organisé comme suit: la section 1 traite la problématique, la section 2 illustre les travaux réalisés pendant l'année en cours. La section 3 présente les perspectives à envisager pour l'année prochaine. En fin la section 4 conclue le rapport.

I. INTRODUCTION ET PROBLÉMATIQUE

La découverte des connaissances à partir de base de données (KDD) est un processus qui consiste à extraire des connaissances cachées à partir de grande masse de données par des méthodes automatiques ou semi-automatiques. C'est aussi un processus itératif et interactif qui est composé de plusieurs phases, notamment le prétraitement et le nettoyage des données. La phase prétraitement est la phase la plus importante dans ce processus, c'est une phase qui occupe 80% d'effort par rapport les autres étapes. Cette phase inclut des opérations comme traitement du bruit, des valeurs aberrantes, les valeurs manquantes, la discrétisation,...etc. La discrétisation à base de l'entropie est utilisée pour minimiser la perte des données. La qualité des résultats d'un processus KDD dépend en grande partie de la qualité des données.

Dans la littérature, il existe plusieurs méthodes de discrétisations qui sont regroupées par plusieurs dimensions telles que : locale vs globale, statique vs dynamique, direct vs incrémentale,...etc. Selon ces catégories il existe plusieurs types de méthodes de mesure de discrétisation (dépendance, précision, entropie).

Parmi ces méthodes, nous distinguons la discrétisation à base de l'entropie est utilisée pour minimiser la perte des données.

Dans cette thèse, nous nous intéressons à étudier les méthodes de la discrétisation des données dans un environnement où les données sont volumineuses et la mise à jour en temps réel avec un degré de certitude plus en moins élevé (Big Data). Notre objectif de recherche est de proposer une approche performante adaptant le processus KDD dans ce contexte.

II. TRAVAUX RÉALISÉS

Nous avons effectué une recherche bibliographique sur la phase prétraitement du processus KDD [8] [9] ainsi le nettoyage des données dans un environnement Big Data [11].

III. PERSPECTIVES

Pour l'année prochaine, nous avons prévus de :

- Compléter l'étude bibliographique sur la phase prétraitement dans le contexte Big Data ;
- Proposer une approche qui permet de discrétiser les données dans le contexte Big Data.
- Finaliser la rédaction d'un article qui traite la correspondance entre les phases du processus KDD et les caractéristiques du Big Data.

IV. CONCLUSION

Après avoir cerné la problématique durant les trois semestres, nous tenterons de la valider par une contribution à soumettre à une revue spécialisée.

V. RÉFÉRENCE

- [1] U.Fayyad, G.Piatetsky-Shapiro, P.Smyth "From Data Mining to Knowledge Discovery in Databases". FALL 1996.
- [2] X.Wu, Fellow, X.Zhu, G.Q.Wu, W.Ding "Data Mining with Big Data". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014.
- [3] ThiagoPoletto, Victor DioghoHeuer de Carvalho, and Ana Paula Cabral Seixas Costa " The Roles of Big Data in the Decision-Support Process: An Empirical Investigation". Springer International Publishing Switzerland 2015 B. delibašić et al. (Eds.): ICDSST 2015, LNBIP 216, pp. 10–21, 2015.
- [4] S.Ramírez-Gallego " Distributed Entropy Minimization Discretizer for Big Data Analysis under Apache Spark". 2015 IEEE Trustcom/BigDataSE/ISPA.
- [5] S.Ramírez-Gallego " Big data preprocessing methods and prospec". Springer International Publishing Switzerland 2015 B. delibašić et al. (Eds.): ICDSST 2015, LNBIP 216, pp. 10–21, 2015.
- [6] SRamírez-Galle " Data discretization taxonomy and big data challenge" 2015 John Wiley & Sons, Ltd
- [7] HUAN LIU " Discretization : An Enabling Technique" Data Mining and Knowledge Discovery, 6, 393–423, 2002
- [8] A. Ragel, B. Cremilleux "MVC—a preprocessing method to deal with missing values". El sevier Knowledge-Based Systems 12 (1999) 285–291
- [9] T. D. Pigott, "A Review of Methods for Missing Data". Educational Research and Evaluation 2001, Vol. 7, No. 4, pp. 353±383
- [10] M.C. Meina Song , J.H. Haihong E "Survey on Data Quality". 2012 World Congress on Information and Communication Technologies

Prise de décision temps réel dans un contexte de Big Data

Yacine Bouhabel

Directeurs de thèse : A. Mostefaoui et M. Koudil

Équipe : (CoDesign)

Résumé— Ce rapport décrit l'état d'avancement de la thèse de doctorat initiée en 2016 sous le thème : Prise de décision temps réel dans un contexte de Big Data.

I. INTRODUCTION

Mon sujet de recherche s'intéresse à la prise de décision temps réel dans un contexte de Big Data. Le sujet étant proposé et encadré par MM. A. Mostefaoui et M. Koudil.

II. PROBLEMATIQUE

Dans un contexte de Big Data où la masse des données dépasse de loin les capacités de gestion développées dans le contexte des SGBDs classiques, de nouveaux modèles de données ont été proposées afin de pouvoir manipuler ces données.

L'objectif initial du sujet consiste à s'intéresser à la phase « Big Data Analytics », et plus précisément la technique de Machine Learning. Il fallait explorer les différents modèles existants et arriver à proposer de nouvelles approches « performantes » permettant l'émission de recommandation à l'utilisateur suite au filtrage des choix et options possibles. Ceci devrait être effectué en temps réel tout en exploitant des données collectées dans le passé.

Après l'exploration du domaine « Big Data Analytics », la problématique de recherche est actuellement plus ciblée vers :

« L'optimisation du traitement dans les moteurs de requêtes Big Data » (Processing Optimization in Big Data Query engines)

III. TRAVAUX REALISES

Plusieurs articles ont été étudiés afin de dégager une piste pour bien définir la problématique de recherche.

Un premier article : « Big Data: A Survey », m'a permis d'identifier les différentes phases du traitement Big Data :

- Génération des données
- Acquisition des données
- Stockage des données
- Analyse des données

Sur d'autres articles, je me suis intéressé aux modèles et algorithmes du Machine Learning utilisés dans la classification des Big Data.

Enfin je me suis intéressé à plusieurs des articles traitant le traitement des requêtes et surtout les moteurs de requêtes (Query Engine).

J'ai aussi mis en place deux machines virtuelles contenant plusieurs outils utilisés dans un environnement Big Data : Cloudera quickstart et Oracle Big Data Lite.

J'ai demandé un stage pour l'année 2017 afin de détailler les caractéristiques et le fonctionnement des moteurs d'exécution des requêtes et de rédiger un article présentant l'état de l'art en la matière.

IV. PERSPECTIVES

- Définition exacte et définitive de la problématique de recherche
- Rédaction d'article couvrant la partie état de l'art
- Etude des optimiseurs de requêtes utilisés dans le contexte de big data.
- Identification de l'apport de la thèse dans le domaine du traitement de requêtes.

Références

- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification.
- Rodrigues, Mário, Maribel Yasmina Santos, and Jorge Bernardino. "Describing and Comparing Big Data Querying Tools." *World Conference on Information Systems and Technologies*. Springer, Cham, 2017.
- Armbrust, Michael, et al. "Spark sql: Relational data processing in spark." *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 2015.

- Soliman, Mohamed A. "Big Data Query Engines." *Handbook of Big Data Technologies*. Springer International Publishing, 2017. 179-217.
- Eltabakh, Mohamed Y. "Data Organization and Curation in Big Data." *Handbook of Big Data Technologies*. Springer International Publishing, 2017. 143-178.

Collecte, Transport et Traitement (CTT) des données de masse dans un environnement mobile et incertain

Fatma Zohra Zahra

Directeur de thèse : Pr Mellouk Abdelhamid

Co-directeur de thèse : Dr Nader Fahima

Date de la première inscription: 27 décembre 2012

Résumé— L'objectif du travail de cette thèse consiste à traiter des données de masse (Big Data) préalablement collectées et transportées d'une façon fiable vers des serveurs. Ces données sont collectées dans un environnement mobile et incertain. Par conséquent, elles sont incertaines et imparfaites. La méthode de traitement adoptée dans ce travail est l'extraction de motifs fréquents qui est une des plus importantes techniques de Data Mining. Les méthodes d'extraction de motifs fréquents proposée dans cette thèse doivent prendre en considération les caractéristiques de Big Data ainsi que le contexte de collecte de données. Les approches proposées sont génériques et peuvent être appliquées sur n'importe quel domaine qui nécessite ce genre de traitement de données. Néanmoins, des problèmes de réseautique tels que la prédiction de la relation QoE/QoS et le monitoring du trafic réseaux seront considérés comme domaine d'application dans cette thèse.

Mots clés— *Extraction de motifs fréquents, Big Data, Données incertaines, Optimisation, Métaheuristiques.*

I. INTRODUCTION

Le traitement des quantités énormes de données fournies par les utilisateurs, et notamment le partage d'informations ubiquitaires (capteurs et senseurs mobiles, caméras, microphones, appareils photos, réseaux de capteurs sans fil, etc.) exige d'une part des procédures avancées de gestion des données pour un stockage et une recherche efficaces d'informations pertinentes, et d'autre part des méthodes raffinées pour extraire et modéliser les connaissances à partir des données [1].

Les techniques d'apprentissage automatique telles que les modèles de Markov, les réseaux de neurones, et les modèles graphiques ont été utilisés avec succès en analyse des données dans les sciences du vivant en raison de leur capacité à manipuler l'aspect aléatoire et l'incertitude. De telles approches, appelées bio-inspirées, ouvrent des pistes de recherche très intéressantes dans le cadre du traitement de plusieurs masses de données. Le travail de

A. Extraction de motifs

Plusieurs axes de recherche algorithmique seront nécessaires dans le cadre de l'extraction de motifs :

- L'optimisation des temps de calcul : Le nombre de sujets et le nombre de variables collectées étant très

important, il conviendra d'optimiser les algorithmes utilisés afin d'améliorer les temps de calcul.

- La qualité et l'intérêt des motifs extraits : Nous cherchons à créer des profils de motifs dont on sait de manière certaine qu'ils auront un intérêt dans le domaine d'application.

B. Assainissement, analyse et comparabilité des données, gestion des incertitudes :

- Dans une approche de mesure de type "crowdsourcing", il est indispensable de traiter la problématique de la validité des données et d'éliminer les éléments aberrants (on parle ici de QoD, quality of data) afin d'être en mesure d'établir des modèles, des recommandations non biaisées par des aberrations statistiques ou d'autres type. Les traitements associés doivent être réalisés en temps réel. De plus, dans une optique de mesure, il est nécessaire de poser les éléments méthodologiques permettant la comparabilité en environnement mobile, l'idée étant de travailler sur une collecte de données relevées à partir d'équipements mobiles.
- La gestion des incertitudes numériques sur un gros volume de données est aussi un élément décisif dans le cadre de cette thèse.

II. TRAVEAUX REALISES

II.1 Collecte et transport des données de masse

La première tâche consiste à faire une étude sur la collecte et le transport des données dans un environnement mobile et de proposer par conséquent une méthode intelligente de transport de ces données. Dans ce cadre, nous avons proposé un framework de collecte et de transport fiable de données pour le suivi à distance des patients via des dispositifs mobiles dans le contexte de Big Data.

II.2 Nettoyage des données de masse

La deuxième tâche porte sur le traitement de la validité des données collectées. On parle ici de la qualité des données (QoD). La QoD est une caractéristique essentielle qui détermine la fiabilité des données pour prendre des décisions. Avec l'avènement de grands volumes de données, la gestion de la qualité des données est plus importante et plus difficile que jamais.

Pour garantir et améliorer la qualité des données, un prétraitement est nécessaire, il consiste à identifier et à éliminer ou à corriger [2] :

- l'inconsistance des données.
- les valeurs manquantes,
- les erreurs et les valeurs extrêmes.

Une synthèse des méthodes et techniques de nettoyage de données existantes pour le traitement de chacun des points cités ci-dessus a été faite, nous avons mis l'accent en premier lieu sur le problème des données incomplètes et leur traitement.

De façon générale, le problème de données incomplètes est présent depuis plusieurs décennies. Ce problème se manifeste lorsque des valeurs n'ont pas pu être observées ou n'étaient pas enregistrées. La présence de ces données entraîne un dysfonctionnement du processus de traitement de données.

Les données manquantes constituent un problème majeur, parce que l'information à disposition est incomplète et donc moins fiable. Le traitement de ces données incomplètes est un problème concret et toujours embarrassant lorsqu'il s'agit de données réelles.

Dans cette optique, nous avons proposé un Framework d'imputation de données manquantes en temps réel basé sur les Réseaux Bayésiens (RBs). En premier temps, un réseau Bayésien est donné par l'expert ou construit à partir des données. Ensuite, nous en déduisons les distributions de probabilité de chaque élément de donnée manquante en utilisant une procédure d'inférence des RBs parallèle et rapide à chaque arrivée de nouvelles données incomplètes. Par conséquent, on obtient toutes les valeurs possibles avec leurs distributions de probabilité correspondantes. La valeur la plus probable sera imputée. L'apprentissage incrémental distribué du modèle de RB garantit l'amélioration et la mise à jour continue du modèle d'imputation de données manquantes. Cette méthode a été implémentée et testée et elle a fait l'objet d'une communication acceptée dans une conférence.

En ce qui concerne le nettoyage des données inconsistantes, une méthode basée sur un algorithme de clustering incrémental et parallèle a été proposée pour détecter la redondance dans les données en temps réel. Néanmoins, il reste à faire des tests plus poussés pour qu'elle puisse faire l'objet d'un article de recherche.

II.3 Extraction de motifs à partir de données incertaines

L'incertitude des données est inhérente dans plusieurs applications émergentes telles que les services de géolocalisation, les systèmes de surveillance à base de capteurs et l'intégration de données. Pour gérer une grande quantité d'informations incertaines, les bases de données incertaines ont été récemment développées [3]. En effet, l'extraction de motifs à partir des données incertaines a attiré l'attention des chercheurs. Ainsi plusieurs travaux sont apparus dans la littérature traitant ce problème, la majorité des solutions proposées se repose sur l'utilisation des modèles

probabilistes pour gérer l'incertitude [3-15]. Il existe un nombre restreint d'algorithmes qui traitent et gèrent l'incertitude dans les masses de données.

Au cours de l'année 2016-2017, il a été envisagé de réaliser les tâches suivantes:

1. Soumission d'un papier intitulé « A Survey on Uncertain Frequent Pattern Mining Algorithms with A Big Data View » au journal ACM COMPUTING SURVEYS, <http://csur.acm.org/>

Dans cet article, nous donnons un aperçu global sur les algorithmes d'extraction de motifs fréquents à partir des données incertaines, en particulier, ceux qui sont proposés pour l'extraction de motifs fréquents dans le contexte de Big Data.

En effet, nous avons explicité les travaux de recherche les plus importants et les algorithmes représentatifs de ce domaine qui traitent d'une part le problème de l'extraction et de la découverte de motifs fréquents à partir de données incertaines et d'autre part, le problème de l'exploitation de motifs fréquents à partir de données incertaines de masse. Un motif fréquent peut être un motif séquentiel, un sous-graphe, un sous-arbre ou un motif spatiotemporel. En outre, un ensemble de données incertaines peut être statique ou dynamique (à savoir les flux de données incertaines). Par conséquent, plusieurs sous-domaines de l'extraction de motifs fréquents à partir de masse de données incertaines ont été étudiés:

- L'extraction des motifs fréquents.
- L'extraction des sous graphes fréquents.
- L'extraction des motifs séquentiels fréquents.
- L'extraction des motifs spatiotemporel fréquents.
- L'extraction des motifs à partir des flux de données incertaines.
- L'extraction des motifs fréquents à partir des données incertaines de masse, sans et avec contraintes.

L'extraction de motifs fréquents à partir des données de masse est un domaine de recherche récent (depuis 2013 à ce jour). L'avènement des données de masse et leur impact dans plusieurs domaines ne cesse de croître; ces données seront de plus en plus diversifiées, gagneront en volume et rapidité. Par conséquent, des solutions efficaces pour traiter les problèmes de Big data dans le contexte de l'extraction des motifs fréquents pourront permettre de suivre la croissance rapide des données en termes de volume. De ce fait, plusieurs verrous de recherche restent ouverts.

2. Soumission d'un papier intitulé « UNCERTAIN FREQUENT ITEMSET MINING BASED ON BEE SWARM OPTIMIZATION » au JOURNAL OF NETWORK AND COMPUTER APPLICATIONS, <https://www.journals.elsevier.com/journal-of-network-and-computer-applications/>

L'extraction de motifs fréquents à partir de masses de données incertaines pose plusieurs problèmes, à savoir:

1. La prise en charge de l'incertitude des données dans la procédure de l'extraction des motifs fréquents.
2. Parmi les caractéristiques de big data, il y a le volume, la variété et la véracité. Le volume et l'espace de recherche des séquences qui sont gigantesques impliquant ainsi un temps d'exécution très important.
3. Le volume des solutions de sortie pouvant être plus grand que les données d'entrée constituent en soi un problème en termes d'interprétation et de pertinence des résultats.

Différentes théories traitant le concept d'incertitude de données ont été proposées, tel que la théorie des probabilités, la théorie des possibilités et la théorie d'évidence. Ces théories ont été exploitées pour gérer l'incertitude des données dans le contexte de l'extraction de motifs fréquents à partir des données incertaines. Le modèle qui a été le plus souvent utilisé est celui du modèle probabiliste. Deux mesures de fréquences ont été proposés, *l'expected support et le probabilistic support*.

En utilisant ce modèle, nous avons proposé une méthode d'extraction de motifs fréquents à partir des données incertaines basée sur une métaheuristique connue, celle de l'optimisation par colonies d'abeilles. D'autres métaheuristicques ont aussi été utilisées, l'optimisation par colonies de fourmis. Le but est de pouvoir établir un comparatif et de chercher de manière empirique la métaheuristique qui s'applique le mieux à ce contexte.

3. Test et validation réelle sur des masses de données incertaines

Concernant le deuxième point relatif au contexte du Big Data. L'extension du travail précédent par la proposition d'un algorithme basé sur le modèle de programmation MapReduce a été réalisée. De plus, un travail sur l'extraction des motifs séquentiels à partir de données incertaines de masse basée sur l'optimisation par colonies d'abeilles a été réalisé et validé par des tests sur une seule machine. « Big Uncertain Frequent sequential Pattern Mining based Bee Swarm Optimization ».

Néanmoins, une validation sur des vraies masses de données et sur un cluster de machines était nécessaire pour finaliser les tests. Des résultats ont été obtenus. L'absence de corpus de bases de données existantes permettant la validation des approches proposées a freiné néanmoins notre avancement. Une demande de séjour au sein du laboratoire de mon directeur de thèse a été formulée et permettra ainsi de disposer in situ de ces masses de données. Le stage m'était accordé en mois de juillet 2017.

4. Théorie d'évidence et données imparfaites

La théorie d'évidence représente aussi un cadre intéressant pour combiner plusieurs types d'incertitude et d'imperfection de données.

Dans le but de prendre en considération l'imperfection de données dans le contexte de l'extraction de motifs fréquents sans passer par un prétraitement et de prendre l'imperfection de données comme un élément enrichissant du résultat, nous avons proposé deux algorithmes d'extraction de données imparfaites.

- L'algorithme **EH-Mine** inspiré de l'algorithme H-mine qui extrait des motifs fréquents à partir des données évidentielles.
- L'algorithme **RS-EApriori** combine rough set theory avec la théorie d'évidence pour l'extraction des motifs fréquents à partir de données imparfaites.

5. Extraction des motifs fréquents sur des dispositifs mobiles

La puissance de calcul croissante des dispositifs mobiles a ouvert de nouveaux axes de recherche en Data Mining. En effet, La mobilité et la portabilité de ces dispositifs imposent implicitement des contraintes de calcul en raison de limitations en énergie, bande passante, affichage, puissance de calcul, et capacité de stockage. Néanmoins, de nombreux algorithmes de data mining ont été exploités avec succès sur les smartphones, on les adaptant aux contraintes de ressources limitées avec les meilleures performances.

Plusieurs travaux ont été proposés dans ce sens, ce qui a engendré un nouveau champ de recherche en Data Mining appelé Mobile Frequent pattern Mining (MFPM). Actuellement, MFDM est basé sur des algorithmes légers qui sont adaptatifs dans des environnements à ressources limitées. Une étude de faisabilité de l'application des méthodes approchées que nous avons proposées dans le cadre de cette thèse dans le contexte mobile a été réalisée.

III. CONCLUSION

Dans le cadre de cette thèse, plusieurs travaux ont été réalisés au cours des années 2015-2017. Une année supplémentaire (réinscription en 6ème année Doctorat) pourra être suffisante pour d'une part, recevoir les avis des papiers soumis aux journaux et d'autre part, valoriser le reste des résultats obtenus par des articles de recherche et la rédaction du manuscrit de la thèse.

Références

- [1] Roberto Tagliaferri, Francesco Masulli, Giorgio Valentini, **Computational intelligence and machine learning in bioinformatics**, Artificial Intelligence in Medicine, Vol.45, (2009), pp. 9-96.
- [2] A. F., Elgamlal, N.A. , Mosa et N.A., Amasha, **Application of Framework for Data Cleaning to Handle Noisy Data in Data Warehouse**.

International Journal of Soft Computing and Engineering (IJSCE), January 05, 2014, pp. Volume-3, Issue-6, pp

226.

[3] Thomas Bernecker · Reynold Cheng, David W, Cheung, Hans- Peter Kriegel, Sau Dan Lee, Matthias Renz, Florian Verhein Liang Wang, Andreas Zuefle, **Model-based probabilistic frequent itemset mining**, Knowl Inf Syst (2013) 37:181–217.

[4] Chun-Kit Chui and Ben Kao, **A Decremental Approach for Mining Frequent Itemsets from Uncertain Data**, PAKDD 2008, LNAI 5012, pp. 64–75, 2008.

[5] Chun-Kit Chui, Ben Kao, and Edward Hung, **Mining Frequent Itemsets from Uncertain Data**, PAKDD 2007, LNAI 4426, pp. 47–58, 2007.

[6] Francisco Guil, Jose M. Juarez, and Roque Marin, **A Possibilistic Approach for Mining Uncertain Temporal Relations from Diagnostic Evolution Databases**, IWINAC 2007, Part I, LNCS 4527, pp. 597–606, 2007.

[7] Toon Calders, Calin Garboni, and Bart Goethals, **Efficient Pattern Mining of Uncertain Data with Sampling**, PAKDD 2010, Part I, LNAI 6118, pp. 480–487, 2010.

[8] Muhammad Muzammal, Rajeev Raman, **On Probabilistic Models for Uncertain Sequential Pattern Mining**, ADMA 2010, Part I, LNCS 6440, pp. 60–72, 2010.

[9] Carson Kai-Sang Leung, Fan Jiang, **Frequent Pattern Mining from Time-Fading Streams of Uncertain Data**, DaWaK 2011, LNCS 6862, pp. 252–264, 2011.

[10] Alfredo Cuzzocrea, Carson K. Leung, **Distributed Mining of Constrained Frequent Sets from Uncertain Data**, ICA3PP 2011, Part I, LNCS 7016, pp. 40–53, 2011.

[11] Carson Kai-Sang Leung, Syed K. Tanbeer, **Fast Tree-Based Mining of Frequent Itemsets from Uncertain Data**, DASFAA 2012, Part I, LNCS 7238, pp. 272–287, 2012.

[12] Fan Jiang, Carson Kai-Sang Leung, **Stream Mining of Frequent Patterns from Delayed Batches of Uncertain Data**, DaWaK 2013, LNCS 8057, pp. 209–221, 2013.

[13] Ying-Ho Liu, **Mining frequent patterns from univariate uncertain data**, Data & Knowledge Engineering 71 (2012) 47–68.

[14] Ying-Ho Liu, Chun-Sheng Wang, **Constrained frequent pattern mining on univariate uncertain data**, The Journal of Systems and Software 86 (2013) 759– 778.

[15] D.S. Rajput, R.S. Thakur, G.S. Thakur, **Fuzzy Association Rule Mining based Frequent Pattern Extraction from Uncertain Data**, IEEE, 2012.

Systeme Virtuel Sécurisé de gestion de clés de chiffrement dans le Cloud-Computing

Doctorant : Saad Fehis

Directeurs de thèse : Dr. Omar Nouali

Équipe : (CERIST / Division Sécurité Informatique)

Résumé—Dans ce travail de recherche nous nous intéressons à l'externalisation des services de type sécurité vers le Cloud-Computing (*La gestion de clés de chiffrement est notre cas d'application*). Ce choix nécessite la création de composants de confiance assurant cette fonctionnalité. En effet, la création de confiance entre le client et le fournisseur de service par la mise en place de mécanismes de sécurité pour la protection de ces composants (*intégrité & interaction*).

I. INTRODUCTION

Dans le cloud computing, les ressources partagées causent de nombreux problèmes de sécurité, faisant que la sécurité est l'une des principaux obstacles à l'adoption des technologies de Cloud-Computing [01]. En plus des menaces traditionnelles, de nouvelles questions devraient être posées telles que: La vulnérabilité due à la virtualisation des infrastructures informatiques [02], l'isolement des données, la gestion de la vie privée et la confidentialité des données. Et tous cela dans un environnement multi-locataire. Sur ce, le vrai challenge, c'est comment créer la confiance (the trust) entre le client et le fournisseur de services dans le Cloud Computing.

Dans ce travail de recherche, on s'intéresse, à offrir aux clients des composants logiciels de confiance (*Plateforme virtuelle de confiance*), qui peuvent assurer la sécurité en tant que service (*Security as a Service, SecaaS*) [20]. Le composant met à la disposition des clients, diverses fonctionnalités de sécurité, à savoir : le filtrage des messages électroniques, filtrage du contenu web, la gestion des vulnérabilités, la gestion des identités, le chiffrement et la gestion des clés de chiffrement (*c'est notre cas d'application*). Ce composant est utilisé en tant que plateforme de confiance d'un système de gestion des clés de chiffrement (*Cryptographic Key Management System CKMS*).

A cet effet, les principaux défis de cette recherche sont :

- (1) La sécurisation du composant ainsi que son fonctionnement dans un environnement totalement virtualisé & multi-locataire (*Cloud-Computing*).
- (2) La création de confiance entre le client et le fournisseur de service, par la mise en place des mécanismes de contrôle et de vérification, par les différentes parties.

Pour atteindre ces objectifs, nous allons présenter dans ce rapport, l'état d'avancement du traitement de ces défis ainsi nos contributions dans cette recherche. Commencant par la description de la problématique dans la section 2, dans la section 3 la démarche suivie dans ce travail de recherche ainsi les travaux réalisés (*Contributions*). Et enfin, une conclusion avec les perspectives.

II. PROBLEMATIQUE :

L'offre CKMS en tant que service, est classée dans la catégorie de la sécurité en tant que service (*Security as a Service SecaaS*) [20]. Ce service, peut être offert sous forme de plateformes partagées ou dédiées (*Compagnies / Group d'utilisateurs*), dépendant de l'architecture adoptée. Ces plateformes sont considérées comme des modules logiciels, assurant leurs fonctionnalités en confiance. Cependant, l'adoption de ces services par les clients, nécessite la création de confiance entre le client et le fournisseur de services, qui est considéré comme le vrai challenge pour l'externalisation ou l'adoption de ce type de plateforme dans le Cloud-Computing.

Le traitement de la problématique de confiance nécessite:

- L'assurance de l'intégrité & le fonctionnement du composant au niveau disque (*Source & exécutable*) et au niveau mémoire (*Exécution*).
- La création des clés dans un contexte de confiance en toute sécurité.
- La protection du dictionnaire de CKMS (*Clés et ses données*) par des mécanismes cryptographiques & de contrôle d'accès.
- La sécurisation de l'interaction & le flux d'information du module dans un environnement multiloculaire (*Partagé avec d'autres logiciels*).
- D'offrir au propriétaire du module ou une tierce de confiance la possibilité de contrôler, de vérifier et d'auditer via des mécanismes de vérification de la chaîne de confiance.

Dans l'objectif de traitement de ces challenges, nous allons présenter dans la section suivante les travaux réalisés.

III. TRAVAUX REALISES :

Afin de traiter les challenges liés à notre problématique, nous avons résumé notre approche par un schéma descriptif (*Voir Fig.1*).

L'approche est dévisée en trois axes, représenté dans le schéma par trois niveaux :

- 1- L'utilisation des TPM (*[4-8]*) en tant que composante hardware, comme (a) Source de confiance pour la création des clés, (b) Un moyen sûr pour l'authenticité des services de CKMS. Donc nous avons proposé d'utiliser le TPM comme base de confiance.

- 2- L'utilisation de la politique de la muraille de chine (*Chinese Wall security policy CWSP [13-21]*) pour créer des murs (*firewall*) autour des instances de CKMS, pour le contrôle des flux d'information & le contrôle d'accès.
- 3- L'utilisation des techniques de chiffrement pour la protection et la manipulation des clés et leurs données (*D-CKMS*) sans déchiffrement.

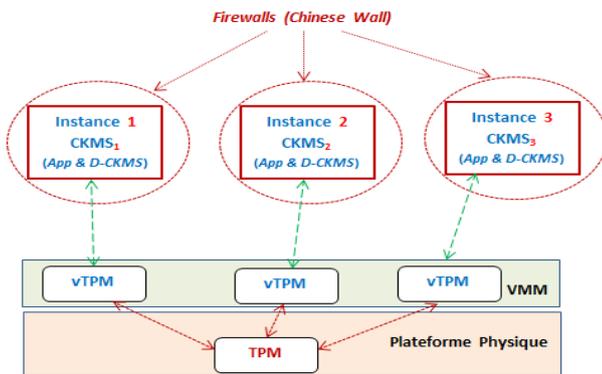


Fig. 1. Schéma de sécurité pour le CKMS-as-a-Service

En effet, nous avons valorisé notre approche par une publication et trois communications internationales orales :

Publication :

- 1- “A New Distributed Chinese Wall Security Policy Model”, *Journal of Digital Forensics, Security and Law (JDFSL)*, Volume 11 | Number 4 (PP 148- 168), 12-31-2016.

Communications:

- 1- “Metadata’s Protection in CKMS-As-A-Security Services”, 4th International Conference on Information Systems and Technologies, Valencia, Spain, 22 – 24 March, 2014.(*Proceedings ICIST 2014*, PP 195- 206 ISBN: 978-0-9561122-5-5)
- 2- “A new Chinese Wall Security Policy Model Based on the Subject’s Wall and Object’s Wall”, First International Conference on Anti-Cybercrime, Riyadh, Saudi Arabia 10 –12 November (*Proceedings ICACC 2015* ISBN: 978-1-4799-7620-1) IEEE Catalog Number: CFP15A69-ART.
- 3- “A Trusted Way for Encryption Key Management in Cloud Computing”, Advanced Information Technology, Services and Systems (*Proceeding AIT2S’2017* ISBN:978-9954-28-542-8), Tangier Morocco, April 14/15, 2017

Enfin, dans le futur proche, nous procéderons à la soumission d’un article pour publication dans une revue scientifique, de titre : «*CKMS as a Trusted Security Service based on TPM in Cloud Computing*»

IV. CONCLUSION & PERSPECTIVES

L’externalisation du CKMS vers une plateforme de confiance dans le Cloud-Computing reste un véritable défi. En effet, la sécurité de toutes les applications (*solutions*) repose

sur la sécurité des clés de chiffrement/signature, qui sont elles-même dépendantes de la sécurité de fonctionnement du CKMS lui-même. C’est dans ce cadre que rentre le challenge de notre travail de recherche et nos contributions.

Comme perspective, nous envisageons de finaliser la rédaction de la thèse et la soutenance.

References

- [1] Cloud Security Alliance. “Top Threats To Cloud Computing”. Technical Report, March 2010. <http://www.cloudsecurityalliance.org/topthreats.html>.
- [2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. “Xen and the Art of Virtualization”. ACM Symposium on Operating Systems Principles (SOSP), 2003.
- [3] National Institute of Standards and Technology; http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf
- [4] TPM Main Part 1 Design Principles Specification Version 1.2 Revision 116 1 March 2011 TCG Published
- [5] Virtualized Trusted Platform Architecture Specification Version 1.0.26 September 27, 2011 TCG PUBLISHED Copyright © TCG 2011
- [6] Krautheim, F.J., Phatak, D.S., Sherman, A.T.: Private Virtual Infrastructure: A Model for Trustworthy Utility Cloud Computing. TR-CS-10-04. University of Maryland Baltimore County, Baltimore, MD (2010), http://www.cisa.umbc.edu/papers/krautheim_tr-cs-10-04.pdf
- [7] F. John Krautheim, D. S. Phatak, A T. Sherman: 211-227 (TRUST 2010: Berlin, Germany) Introducing the Trusted Virtual Environment Module: A New Mechanism for Rooting Trust in Cloud Computing.
- [8] D. Chang, X. Chu, Y. Qin, D. Feng : “TSD Trusted Service Domain: A Flexible Root of Trust for the Cloud; (2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications)
- [9] NIST DRAFT Special Publication 800-130 April 2012 “A Framework for Designing Cryptographic Key Management Systems
- [10] David D. C. Brewer and Michael J. Nash: The Chinese Wall Security Policy, in Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA: IEEE Press, May, 1988, pp 206-214.
- [11] T. Y. Lin, “Chinese Wall Security Policy–An Aggressive Model”, Proceedings of the Fifth Aerospace Computer Security Application Conference, December 4-8, 1989, pp. 286-293.
- [12] Tsau Young Lin: Chinese Wall Security Model and Conflict Analysis. COMPSAC-2000: 122-127.
- [13] Tsau Young Lin: Placing the Chinese Walls on the Boundary of Conflicts - Analysis of Symmetric Binary Relations. COMPSAC 2002: 966-974
- [14] Tsau Young Lin: Chinese Wall Security Policy Models: Information Flows and Confining Trojan Horses. DBSec 2003: 275-287.
- [15] Tsau Young Lin: Chinese wall security policy-revisited a short proof. SMC 2007: 3027-3028
- [16] R. Sandhu. Lattice-Based Enforcement of Chinese Walls. Computers & Security, 11(8):753–763, 1992.
- [17] D. E. Bell and L. J. LaPadula. Secure Computer System: Unified Exposition and MULTICS Interpretation. Technical Report MTR-2997 Rev. 1 and ESD-TR-75-306, rev. 1, The MITRE Corporation, Bedford, MA 01730, Mar. 1976.
- [18] Alireza Sharifi, Mahesh V. Tripunitara: Least-restrictive enforcement of the Chinese wall security policy. SACMAT 2013: 61-72
- [19] “A new Chinese Wall Security Policy Model Based On the Subject’s Wall and Object’s Wall”, First International Conference on Anti-Cybercrime, Riyadh, Saudi Arabia 10 –12 November (Proceedings ICACC 2015 ISBN: 978-1-4799-7620-1) IEEE Catalog Number: CFP15A69-ART.
- [20] A. Jerry, B. Alan, B. Alan, C. Dave, P. Nils, K. Paul, and R. Jim, “Defined categories of service 2011,” Cloud Security Alliance, Security as a service working group, 2011. [Online]. Available: <http://www.cloudsecurityalliance.org/guidance>

Contributions à la détection et à l'agrégation des événements dans les microblogs

Meghatria Riadh

Directeurs de thèse : Pr Latiri Chiraz, Pr Nader Fahima

Équipe :

Abstract—La tendance actuel à tout publier sur les microblogs, notamment Twitter, rend ces derniers une source importante d'information, pouvant être exploitée dans plusieurs domaines. Toutefois, extraire efficacement les informations pertinentes émergées dans un contenu aussi grand, plein de répétitions et d'informations inutiles représente un vrai challenge. Dans ce cadre de recherche s'inscrit notre sujet de thèse qui s'intéresse à la détection des événements dans les microblogs.

Mots clés : Fouille de réseaux sociaux, fouille de textes, twitter, détection d'événements, classification, Analyse Formelle de Concepts, agrégation.

I. INTRODUCTION

Ces dernières années, le développement du web social et collaboratif 2.0 a rendu les internautes plus actifs au sein des réseaux participatifs. Ces derniers (Twitter, facebook, Youtube, ...) servent d'un bon moyen de communication pour les gens qui veulent partager et échanger des informations sur différents événements du monde réel allant de ceux largement ou mondialement connus (concert d'une star de musique) aux événements locaux (accident, manifestation, regroupement social local). Les messages courts publiés sur les microblogs tel que Twitter peuvent généralement refléter ces événements dès leurs occurrence. Ces nouvelles formes de publication s'inscrivent désormais dans une logique d'intelligence collective et de gestion des connaissances, et ont un potentiel inattendu en termes de veille stratégique et d'analyse d'opinions.

En effet, les professionnels de l'information peuvent les utiliser comme nouvelles ressources pour y rechercher de l'information. Or, ces derniers se confrontent à l'abondance de ce contenu informationnel. Comment effectuer un tri efficace à partir de cette masse de ressources, pour ne conserver que les informations pertinentes en fonction d'une problématique ? La problématique de la thèse essaye de répondre à cette question.

II. PROBLÉMATIQUE

Plusieurs travaux de recherche concernant l'extraction des informations pertinentes à partir des réseaux sociaux et des microblogs sont entrepris et ce pour un large éventail d'application. Dans cette thèse nous nous intéressons particulièrement à la problématique de détection d'événements à partir des microblogs [1] tel que Twitter. L'objectif derrière est de pouvoir analyser les événements qui intéressent le plus les utilisateurs de Twitter ou encore quantifier l'impact des événements sur les utilisateurs de Twitter [4]. Toutefois, cette tâche demeure non-triviale, sans oublier que les tweets rapportant les événements importants sont noyés par un grand nombre de tweets marginaux.

Dans ce contexte, et après une première étude bibliographique portant sur la détection des événements à partir des microblogs, les méthodes de classification/clustering à base d'Analyse Formelle de Concepts et les approches de fouille de textes à base d'AFC, les contributions attendues dans le cadre de cette problématique, s'articulent en trois parties, à savoir :

- 1) Déterminer à partir des données envoyées sur les microblogs (tweets), les événements qui stimulent l'intérêt des utilisateurs durant une période donnée et les dates saillantes de chaque événement. Nous supposons qu'un événement est représenté par plusieurs termes dont la fréquence évolue en fonction des spécificités des moments durant la période analysée. Pour ce faire, nous proposons des approches de fouille de textes et de classification à base d'Analyse Formelle de Concepts. Ce couplage permettra de décliner la contribution selon deux axes : (1) identifier des regroupements événements inconnus à partir d'un ensemble de tweets (concepts fermés fréquents) et gérer la redondance; (2) classer des tweets selon le type d'événement.
- 2) Une fois les événements identifiés, l'idée est de développer un modèle d'agrégation dans l'objectif de regrouper les tweets les plus pertinents décrivant un événement donné.
- 3) Pour l'évaluation expérimentale de nos méthodes, nous projetons de mener le protocole d'évaluation sur les collections de TREC Microblogs 2015 Temporal Summarization Track.

III. TRAVAUX RÉALISÉS

Depuis la première inscription en mois d'octobre 2016, l'acceptation en mois de février 2017 (point de commencement), nous essayons toujours de dresser un état de l'art sur la détection des événements à partir des microblogs, en lisant les articles et les thèses traitants sur le sujet, les résumant et faisant la synthèse de chaque travail (points faibles et points forts). Cette première étape d'étude bibliographique, importante qu'elle soit, n'a pas pu se dérouler d'une manière continue (sans interruptions), ce qui nous a causé du retard, pour cette première année, et qui devrait se rattraper durant l'année en cours.

À partir des articles lus, nous pouvons noter que Twitter est le microblogue de choix, pour les recherches scientifiques qui s'intéressent à tirer des informations à partir des médias sociaux, prélever les intentions ou même les émotions des utilisateurs et identifier les événements qui stimulent leurs intérêts à un moment donné. Ceci est dû à la popularité de Twitter, son haut taux d'utilisation et principalement à

l'accessibilité public de son contenu via des interfaces de programmations (APIs). Streaming API et REST API sont les interfaces plus connues et plus utilisées de Twitter. La première permet un accès temps réel à tous le flux de données circulant sur Twitter (c.à.d tous les tweets du moment), tandis que la deuxième fournit des données historiques convenant à des recherches pointées (profile d'un utilisateur, ses tweets...).

Le contenu d'un tweet, limité à 140 caractères, peut être un avis, une information ou un témoignage. Dans un tweet, il est possible de mentionner un utilisateur ou lui adresser une réponse en plaçant un '@' devant son nom d'utilisateur (@username). L'utilisation de 'RT@username' permet de faire un retweet (c.à.d transférer un tweet tout en mentionnant son auteur). Dans un tweet, les sujets peuvent être étiquetés avec un mot hashtag, un mot précédé par un dièse # (hash en anglais). En cliquant sur un hashtag, la liste des tweets ayant le même hashtag s'affiche.

A. Détection des évènements

La détection des évènements a longuement été traitée dans le programme TDT (Topic Detection and Tracking) qui est concerné par l'organisation événementielle du flux de documents, et ce dans le but de fournir des outils technologiques pour le suivi des actualités provenant des différentes sources de médias traditionnels. La détection des évènements est composée de trois phases : prétraitement (Stopwords filtering, Stemming, ...), représentation des données (term vector, ...) et organisation des données (clustering).

Dans notre thèse, nous nous intéressons aux média actuels (microblogues). La détection d'évènements à partir de Twitter est un domaine de recherche dynamique qui s'appuie sur des techniques de divers domaines: l'apprentissage automatique, le traitement automatique de la langue, le Datamining et la fouille de texte.

En effet, identifier un événement inconnu ou connu a priori revient à détecter une thématique *saillante* : étant donné un intervalle I , une thématique dite saillante si elle a attiré un niveau d'attention beaucoup plus élevé durant I que pendant le reste de la période d'observation. Une thématique est définie généralement par un mot principal et un ensemble pondéré de mots liés. Ainsi, un événement est défini comme une thématique saillante et une valeur reflétant la magnitude d'impact sur les utilisateurs de Twitter.

Dans la littérature, nous distinguons trois principales familles de méthodes pour la détection d'évènements à partir de microblogues, à savoir : *i*) des méthodes basées sur la *fréquence des mots* [2]; *ii*) des méthodes à base de *topic-models* [3], [5]; et des méthodes à base de *clustering de mots/n-grams de mots* [6]. Il a été constaté que la quasi-majorité des méthodes se concentrent sur le contenu textuel, ignorant l'aspect social (popularité de l'auteur, attribut spatio-temporel, etc). également, la durée des évènements est fixée à l'avance (typiquement une journée). Les différentes techniques proposées pour la détection d'évènements à partir de Twitter peuvent aussi être classées selon le *type d'évènement* (spécifié ou non-spécifié), *l'application cible* (RED: détection rétrospective d'évènements, ou NED: détection des nouveaux évènements) et *la méthode de détection* (supervisée, non supervisée ou hybride).

IV. PERSPECTIVES

- Terminer l'étude bibliographique;
- Rédiger un état de l'art portant sur la détection des évènements à partir des microblogues, en mettant en avant les approches de classification et de fouilles de textes impliquées;
- Proposer et implémenter une méthode pour identifier des regroupements d'évènements inconnus à partir d'un ensemble de Tweets.

V. CONCLUSION

Après avoir développé les points mentionnés dans la section précédente, nous ferons une étape de test, évaluation et discussion des différents résultats. Une fois les évènements identifiés, nous poursuivrons les contributions par la proposition d'un ordonnancement par apprentissage des tweets pertinents dans le but de générer un *contexte* décrivant avec efficacité l'évènement identifié. Nous utiliserons les collections de TREC Microblogs 2015 pour nos tests expérimentaux.

Dans le présent rapport, nous avons résumé le déroulement de la thèse : sa problématique, ses objectifs ainsi que les travaux réalisés jusque-là. Enfin, nous avons cité les perspectives fixées pour l'année qui suit.

REFERENCES

- [1] Farzindar Atefeh and Wael Khreich. A survey of techniques for event detection in twitter. *Comput. Intell.*, 31(1):132–164, February 2015.
- [2] Housseem Eddine Dridi and Guy Lapalme. Détection d'évènements à partir de twitter. *TAL*, 54(3):17–39, 2013.
- [3] Wei Gao, Peng Li, and Kareem Darwish. Joint topic modeling for event summarization across news and social media streams. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12*, pages 1173–1182, New York, NY, USA, 2012. ACM.
- [4] Adrien Guille and Cécile Favre. Event detection, tracking, and visualization in twitter: a mention-anomaly-based approach. *CoRR*, abs/1505.05657, 2015.
- [5] Nathan Keane, Connie Yee, and Liang Zhou. Using topic modeling and similarity thresholds to detect events. In *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 34–42, Denver, Colorado, June 2015. Association for Computational Linguistics.
- [6] Jianshu Weng and Bu-Sung Lee. Event detection in twitter. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*, 2011.

Approche de sécurisation des communications dans l'internet des objets (IoT)

Abbadi Malika

Directeurs de thèse : Koudil Mouloud, Professeur à ESI

Équipe : CoDesign

Abstract—Ce papier présente l'état d'avancement des travaux de recherche dans le cadre de la thèse de doctorat avec une réorientation vers le domaine de la qualification des données pour l'internet des objets.

I. INTRODUCTION

On travaillant sur la sécurité des communications dans l'internet des objets, les premiers travaux étaient sur les protocoles et standards de sécurité adaptés depuis l'internet pour l'IoT[1][2][3]. un autre courant s'est intéressé au développement d'infrastructures adéquates pour la sécurité dans l'IoT, comme la technologie 5G. Une autre approche est d'établir des stratégies de confiance pour garantir la sécurité des transmissions dans un environnement internet des objets[5]. Ces stratégies sont généralement basées sur la confiance de la source de données transmises[4][6][7]. avec un besoin pour valider les données par comparaison avec des données transmises par d'autres sources. Cette validation des données ne peut être développée sans explorer la qualité des données.

Les recherches menées cette année, se sont orientées essentiellement vers les travaux liés à la qualité de données.

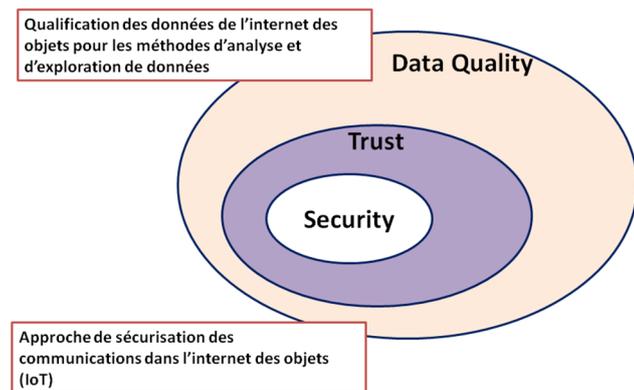


Fig 1. position de la qualité des données dans l'internet des objets

La qualité des données est un domaine dont les recherches se multiplient depuis des décennies. Appliqué à plusieurs domaines comme les bases de données, les données Web, les données des réseaux sociaux et les réseaux de capteurs. Récemment, émergent des solutions pour l'internet des objets, mais l'alliance entre le contexte de l'internet des objets et la qualité des données reste un domaine très riche et peu exploré.

II. PROBLÉMATIQUE

Avoir des infrastructures et des canaux de communications fiables pour l'internet des objets doit être renforcé par des données exactes et précises. Définir, mesurer, évaluer et améliorer la qualité des données sont les parties principales pour avoir des communications fiables via l'internet des objets. valorisation des données dans un environnement de l'internet des objets comprend plusieurs phases passant de la collecte des données, le prétraitement, les méthodes de stockage et les méthodes d'analyse.

La communauté de recherche s'est concentrée sur les méthodes de nettoyage de données notamment la détection d'anomalies et la déduction de données manquantes. Karkouch et al en 2016 [9] parcourt les travaux effectués sur ce volet dans les réseaux de capteurs sans fil et les systèmes RFID étant les deux technologies dominantes de l'internet des objets. Le nettoyage des données tout en étant incontournable, est rudimentaire et insuffisant. L'internet des objets, caractérisée par l'hétérogénéité des sources de données force la recherche de solutions et méthodes adaptées de traitement de la qualité des données. D'autres travaux allant des premiers travaux de référence de Wang et Strong en 1996 [8] jusqu'aux derniers travaux de Karkouch et al (2017)[10] et Hoyos (2016)[11] proposant des solutions faisant abstraction de la plateforme de travail de destination. C'est seulement après la définition des exigences de l'utilisateur que le code et les bases de données sont générés.

D'autres chercheurs se sont attaqués aux problèmes de sélection de sources fiables de données pour la collecte des données. Peu de travaux se sont intéressés à la sélection des données (datasets). On peut nommer les travaux de Hong et Huang (2017)[12] sur la sélection d'ensembles de données GIS pour assurer qu'on a toutes les données nécessaires à l'analyse (completeness DQ dimension). D'après notre étude, plusieurs problèmes restent à traiter:

- 1) Les travaux existants ne prennent pas en considération l'hétérogénéité des données, ils considèrent des données de même format.
- 2) Les travaux existants ne considèrent pas les différences dans le format d'écriture des meta-data/annotations (des représentations spécifiques). Alors, ces annotations seront inexploitables.
- 3) Plusieurs solutions sont liées au domaine et à des dimensions de qualité spécifiques. Ces solutions sont conçues selon les caractéristiques d'une dimension de qualité.
- 4) Les solutions ne bénéficient pas des données précédentes (historic data).

- 5) Les solutions utilisent des méthodes de calcul de qualité liées aux caractéristiques des réseaux et du matériel déployés (ex: le débit, capteur).

L'objectif est de proposer des solutions pour Qualifier les données de l'internet des objets pour les méthodes d'analyse et d'exploration de données. Ces solutions doivent satisfaire les besoins de:

- 1) Qualifier les données après l'étape de prétraitement (après le nettoyage des données s'il existe).
- 2) Réduire le volume de données à traiter.
- 3) Sélectionner des ensembles de données (Datasets) en accordance avec les exigences de l'utilisateur.

III. TRAVAUX RÉALISÉS

Une étude de la littérature sur la qualité des données d'un côté et les caractéristiques des données dans l'internet des objets et le Big Data, a ressortie une classification des travaux selon plusieurs critères. Sur la base de cette classification, on a pu préciser la problématique de la recherche détaillée en section 2. Les travaux ont été classés en plusieurs catégories selon qu'ils traitent la qualité par phase de traitement des données (collecte de données, nettoyage de données, sélection de sources ou de données), par dimension de qualité (accuracy, timeliness, consistency, completeness) et que les travaux soient des solutions génériques, des solutions liées aux dimensions de qualité ou des solutions qui proposent des compromis entre la qualité des données et d'autres critères qui influencent les transmissions comme l'énergie.

Cette classification détaillée a permis de relever plusieurs pistes intéressantes. Notamment, plusieurs dimensions de qualité comme la réputation et la sécurité d'accès, restent non explorées dans le contexte de l'internet des objets. Une étude de l'importance de ces dimensions et de leurs impacts pourrait faire l'objet d'un article.

Une solution sur laquelle on travaille est de proposer à l'utilisateur analyste, des données brutes, des données traitées, et des valeurs de qualité des données disponibles. Ces valeurs seront générées et sauvegardées automatiquement en forme de méta data ou d'annotations. Ces informations supplémentaires sur les données seront accompagnées de détails sur les méthodes de calcul pour plus de crédibilité.

IV. PERSPECTIVES

L'objectif est une solution qui permet de qualifier (donner un poids) automatiquement aux paramètres liés à la donnée (caractéristiques). Ces informations supplémentaires déterminent et contribuent à la validité de la donnée pour l'utilisation en cours (est-ce qu'elle est importante pour nous, est-ce qu'on peut faire confiance). Le poids attribué n'est pas fixe mais plutôt variable selon l'utilisation. Aussi, on a besoin de définir les relations entre ces paramètres par une fonction coût. Cette solution se fera à l'aide des méthodes de Machine Learning pour un apprentissage automatique de l'importance et l'utilisation des paramètres liés à la donnée. Cette approche doit être étudiée sur deux niveaux: la sélection de type de données (attributs comme température, humidité) et la sélection et la qualification dynamique des caractéristiques des données comme le degré de précision du relevé de température et est-ce qu'on a des

informations sur la mesure employée (ex: degré Fahrenheit ou degré celsius). les applications et outils utilisés actuellement pour l'analyse des données comme a des téraoctets de données fournies continuellement par l'internet des objets.

V. CONCLUSION

le domaine sur lequel on travaille, qualifier des données de l'internet des objets pour les méthodes d'analyse, est un domaine nouveau dont la communauté de recherche s'y intéresse de plus en plus. Les publications sur ce sujet apparaissent continuellement répondant à des questions et en posant de nouvelles.

REFERENCES

- [1] S. Raza, S. Duquennoy, J.H. Oglund, U. Roedig, and T. Voigt, Secure communication for the internet of things ? A comparison of link-layer security and IPsec for 6LoWPAN, SECURITY AND COMMUNICATION NETWORK, Security Comm. Networks 2011; 00:1?14 DOI: 10.1002/sec.
- [2] S. Sicari, A. Rizzardi, L.A. Grieco and A. Coen-Porisini Security, privacy and trust in internet of things: a road ahead, Elsevier Computer networks 76 (2015) 146-164, DOI: 10.1016/j.comnet.2014.11.008.
- [3] Yosra Ben Saïed, Collaborative security for the internet of things, Institut National des Télécommunications, 2013.; NNT : 2013TELE0013; ;tel-00879790.
- [4] N.B. Truong, T.W. Um and G.M. Lee, A reputation and knowledge based trust service platform for trustworthy social internet of things, 19th International ICIN Conference - Innovations in Clouds, Internet and Networks - March 1-3, 2016, Paris.
- [5] M. Nitti, R. Girau, L. Atzori, A. Iera, and G. Morabito, A subjective model for trustworthiness evaluation in the social internet of things, 23rd Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, conference paper, DOI:10.1109/PIMRC.2012.6362662.
- [6] G. Lizet, W. Jingpe, and S. Bin, Trust Management Mechanism for Internet of Things, China Communications? February 2014, pp 148-156.
- [7] Y.B. Saïed, A. Olivereau, D. Zeghlache and M. Laurent Trust management system design for the internet of things: a context-aware and multi-service approach, Elsevier Computer and security XXX (2013) 1-15, DOI.org/10.1016/j.cose.2013.09.001.
- [8] R. Wang, D. Strong, Beyond accuracy: what data quality means to data consumers, Published in: Journal of Management Information Systems archive, Volume 12 Issue 4, March 1996, Pages 5-33, doi:10.1080/07421222.1996.11518099.
- [9] A. Karkouch, H. Mousannif, H. Al Moatassime, T. Noel, Data quality in internet of things: A state-of-the-art survey, Journal of Network and Computer Applications 73 (2016)57-81, 1084-8045/ 2016 Elsevier, http://dx.doi.org/10.1016/j.jnca.2016.08.002.
- [10] A. Karkouch, H. Mousannif, H. Al Moatassime, T. Noel, A model-driven framework for data quality management in the Internet of Things, J Ambient Intell Human Comput, © Springer-Verlag Berlin Heidelberg 2017, DOI 10.1007/s12652-017-0498-0.
- [11] J.R. Hoyos, J. Garcia-Molina, J.A. Botia, D. Preuvneers, A model-driven approach for quality of context in pervasive systems, Computers and Electrical Engineering 0 0 0 (2016) 1?20, http://dx.doi.org/10.1016/j.compeleceng.2016.07.002
- [12] J.H. Hong, M.L. Huang, Enabling smart data selection based on data completeness measures: A quality-aware approach, International journal of Geographical Information science, doi: 10.1080/13658816.2017.1290251.

Development of Multiobjective application mapping for heterogeneous network on chip

Bougherara Maamar

Directeurs de thèse : Koudil Mouloud and Nedjah Nadia

Équipe : Network on chip

Abstract—Network-on-chip (NoC) is considered the next generation of communication in embedded system. In this case, an application is implemented by a set of collaborative intellectual propriety blocks (IPs). The selection of the most suited block from a library of IPs is an NP-complete problem. In this purpose, MOEA (multiobjective evolutionary algorithm) are the most suitable to yield the best selection of IP to implement efficiently a given application on a NoC infrastructure.

I. INTRODUCTION

As the integration of semiconductor increase more complex cores for a system on chip are launched a complex System-on-Chip (SoC) is formed by interconnected heterogeneous component, the interconnection is formed by NoC. A NoC is similar as a general network but with limited resources, power and area. Each component of the NoC is an intellectual propriety IP block that can be a general or special purpose such as a processor, memory and DSP [1]. A NoC is designed to run a specific application. This kind of applications is generally limited by the number of the tasks that are implemented onto IP block. An IP Block can implement more than a single task of the application. It can execute many tasks as a general-purpose processor does. In contrast, for instance a multiplier IP block for floating point number can only multiply floating point numbers. The number of IP blocks is growing up fast. So, it is necessary to choose the adequate IP block before as to yield an efficient NoC-based design for any application. The choice becomes harder when the number of task increases. it is also necessary to map these block onto the NoC available infrastructure, which consists of set of cores communicating using switches. Different optimization criteria can be pursued depending on information details about the application and IP blocks. The application is a graph of task called Task Graph (TG). The features of an IP block can be determined from its companion library. The task assigned and IP block mapping are key search for an efficient NoC-based designs [2]. These two problems can be solved using multi-objective optimization. Some of the objectives are conflicting because of their nature. IP assignment and IP mapping are classified as NP-hard problems [3]. The deterministic techniques are not viable to solve such problems. So it is mandatory to use Multi- Objective Evolutionary Algorithms (MOEAs) with specific objective functions. For this purpose, one needs to select the best minimal set of objectives to be optimized. The wrong set of optimized objectives can lead to average instead of best results.

II. IP ASSIGNMENT AND MAPPING PROBLEMS

The platform-based design methodology for SoC encourages the reuse of components to reduce costs and to reduce

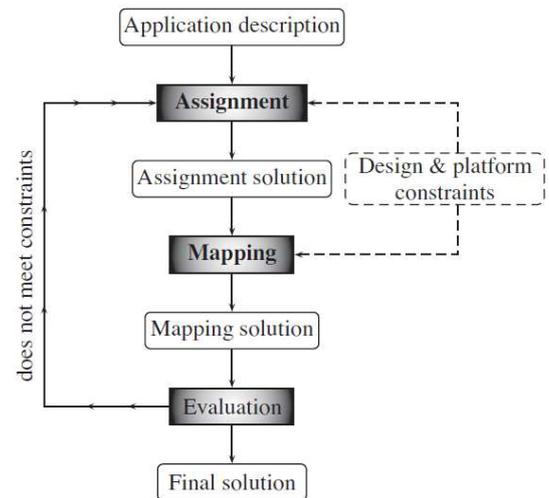


Fig. 1. Platform-based NOC Design Workflow

the time-to-market of new designs. The designer of NoC-based systems faces two main problems: selecting the adequate set of IPs and finding the best physical mapping of these IPs into the NoC structure. On a platform-based design, the selection of IPs is called IP assignment stage and the physical mapping is called IP mapping stage. A typical platform-based NoC design work-flow is shown in Fig 1.

A. The IP Assignment Problem

IP Assignment is the first step before mapping the application onto NoC [4]. The objective of IP Assignment is to select, from an IP library (IP repository), a set of IPs that exploit re-usability and optimize the implementation of a given application in terms of time, power and area requirements. During this step, no information about physical location of IPs onto NoC is given. The optimization process must be done based on TG and IP features only. The result of this step is a set of IPs that should maximize the NoC performance, *i.e.* minimize power consumption, hardware resources as well as the total time execution of the application. Recall that the result of this step is produced in the form of an ACG for the application's task graph, wherein each task has an IP associated with it.

Note that the number of possible assignments is defined as

in Eq 1:

$$\#A = n_1 \times n_2 \times \dots \times n_{m-1} \times n_m, \quad (1)$$

wherein m represents the number of tasks used in the application and n_i the number of IPs that can be assigned to task i .

B. The IP Mapping Problem

Given an application CG, the problem that we are concerned with here is to determine how to topologically map the selected IPs onto the network structure, such that the objectives of interest are optimized. At this stage, a more accurate evaluation can be done taking into account the distance between resources and the number of switches and channels crossed by a data package during a communication session. The result of this process should be an optimal allocation of the prescribed IP assignment, to execute the application on the NoC.

The mapping stage uses the result obtained from the assignment, which consists of many non-dominated solutions. Let s be the number of distinct assignments evolved and p_i be the number of processors used in assignment i and n_i be the minimal number of resources in the NoC to be utilized in the implementation of the application with assignment solution i . In this case, the total number of possible mappings is defined by the sum of all the mappings corresponding to each one of the evolved assignments as described in Eq 2:

$$M_S = \sum_{i=1}^S \frac{n_i!}{(n_i - p_i)!} \quad (2)$$

III. WORK COMPLETED

Before considering a contribution, a bibliographical study was carried out. A new classification of NoC Application mapping is proposed according to the many classes. A summary of the techniques used in this field The proposed survey highlighting more the multi-objective techniques and we classified these techniques according to:

A. strategy used

The strategy used in multi-objective application mapping use several objective functions taken into account, two multi-objective methods can be used agregation methods and pareto methods

B. Algorithms used

MOEAs(multi-objective evolutionary algorithms) are the most used algorithms in multiobjective problems and they are classified by two class elitist and non-elitist MOEAs, in elitist algorithms the best solutions are saved in an archive, using an archive improve the solution to converge to the global optimal solutions

C. objective used

Several objectives are taken into account, these objectives have also been classified according to 3 classes Computing Objectives: they are objective when executing the application (time, space, energy). communication objectives are calculated when communicating between IPs using a routing algorithm. reliability are the objectives to see fault tolerance of the NoC. Since there is no work that presents multi objective methods, this work takes into account all the multiobjective methods that exist classifying according to our classification. Once this work is finalized and validated, it can be submitted to a journal.

IV. PERSPECTIVE

a contribution is in preparation takes into account the two steps of the mapping problem (assignment step and the mapping step) most of the works take both steps as a single problem and propose a single approach. We think that the problem can be divided into two difficult problem, and also propose different approaches to each of them.

V. CONCLUSION

The task assigned and IP block mapping are key search for an efficient NoC-based designs. These two problems can be solved using multi-objective optimization. Some of the objectives are conflicting because of their nature. IP assignment and IP mapping are classified as NP-hard problems. Our PhD consists in proposing a mapping method taking into account the two phases of the problem.

REFERENCES

- [1] Jingcao Hu and Radu Marculescu. Energy-aware mapping for tile-based NoC architectures under performance constraints. In ASPDAC: Proceedings of the 2003 conference on Asia South Pacific design automation, pages 233–239, New York, NY, USA, 2003. ACM.
- [2] Umit Y. Ogras, Jingcao Hu, and Radu Marculescu. Key research problems in NoC design: a holistic perspective. In Proceedings of the 3rd IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, CODES+ISSS 2005, Jersey City, NJ, USA, September 19–21, 2005, pages 69–74. ACM, 2005.
- [3] M. R. Garey and D. S. Johnson. Computers and intractability; a guide to the theory of NP-completeness. Freeman and Company, 1979.
- [4] Nadia Nedjah, Marcus Vincius Carvalho da Silva, Luiza de Macedo Mourelle: Preference-based multi-objective evolutionary algorithms for power-aware application mapping on NoC platforms. Expert Syst. Appl. 39(3):2771–2782 (2012)

Une Ontologie de domaine pour l'aide à la décision dans un cadre juridique

Keltoum Benlaharche

Directeurs de thèse : Djamel Meslati & Nabila Nouaouria

Résumé—Dans le présent travail, nous nous intéressons à la construction d'une ontologie d'aide à la décision appliquée dans un domaine particulier de la législation islamique, en l'occurrence, les transactions bancaires islamiques. En effet, il s'agit là d'un domaine où les questions sont fréquentes et la recherche de réponse reste dans la majeure partie du temps basée sur l'expertise humaine uniquement. Cette ontologie servira par la suite pour l'inférence dans ce domaine à fin de supporter le législateur (El Mufti) dans sa démarche de génération de fatwa et de construction de l'argumentation accompagnant El fatwa (El-Isnad). Il s'agit là d'une inférence dans un style de raisonnement basé cas interprétatif. Le système expert qui en découlera fonctionnera dans un mode inspiré de la manière naturelle des Muftis, en l'occurrence, l'analogie (El-Kiass).

I. INTRODUCTION

En législation islamique, El Fatwa est un avis juridique donné par un spécialiste de la loi islamique sur une question particulière. El Fatwa consiste en l'exposition d'un avis juridique en accord avec les préceptes islamiques[1].

Nous proposons un système expert basé ontologie pour générer des Fatwas, en réutilisant les fatwas des anciennes situations, par imitation du Mufti. Nous avons limité le domaine d'El fatwa au domaine des transactions bancaires islamiques.

II. PROBLEMATIQUE

La problématique abordée portera sur un domaine d'application peu abordé qui est le domaine de la législation islamique ; les transactions bancaires ; et procèdera par construction d'une ontologie pour un système expert basés cas.

Notre objectif est de créer une base de connaissances qui représente le domaine de la finance islamique qui aidera le législateur à prendre des décisions sur une question posée dans le domaine en se basant sur les situations similaires déjà résolues.

III. TRAVAUX REALISES

En 2004, un projet de recherche a été lancé par le laboratoire LISCO de l'université Badji Mokhtar de Annaba. Dans l'objectif de construire un système expert d'aide à la fatwa. Ce système a vu le jour en 2006 [2] un outil intelligent pour assister el Mufti dans sa démarche de générer El Fatwa. Cet outil a été limité uniquement au domaine des aliments et boissons dans le but d'illustrer le raisonnement. En 2014, une nouvelle architecture du système basés cas a été proposée [3]. Notre travail est inspiré sur cette dernière référence. En vue de présenter des améliorations, nous avons introduit l'aspect ontologique et appliqué le travail au domaine des transactions bancaires.

En 2011, Mamadolimova, Ambiah et Lukose (dans [4]) ont modélisé les connaissances (concepts et propriétés) associées à la finance islamique en utilisant une ontologie. Cette ontologie peut être utilisée par la suite pour les systèmes d'information dans la vérification et dans d'autres types d'activités de gestion des risques. Le modèle en spirale est la technique utilisées pour modéliser l'ontologie. 'Topbraid composer' est l'outil utilisé pour l'ingénierie de l'ontologie. c'est l'unique ontologie trouvé qui porte sur le domaine de la finance islamique.

L'étude menée en 2009 par M. Gruninger et Fox (voir [5]) présente une approche pour le développement d'un système expert d'évaluation de la qualité financière d'une entreprise. Ils ont appliqué une ontologie pour modéliser le contenu du domaine de connaissance. Ils ont utilisé des règles de décision pour représenter la connaissance opérationnelle. Le système a été intégré en utilisant Protégé pour le domaine de connaissance et JESS pour la base de connaissance opérationnelle dans un seul système expert complet. La méthodologie utilisée pour construire l'ontologie est TOVE (Toronto Virtual Entreprise). L'ontologie est constituée des classes, des relations entre ces classes, des attributs et des individus qui sont le domaine de connaissance (taxonomie).

Une autre étude menée en 2014 par Rajput et al. (voir [6]) propose une ontologie basée système expert pour détecter les transactions suspectes dans les institutions financières. L'utilisation de l'ontologie rend les systèmes experts plus efficaces, requière peu de calcul et permet de réutiliser la base de connaissance dans des différentes applications de domaines similaires. Ils ont développé un mécanisme permettant de prévenir les activités de blanchiment d'argent. L'ontologie est composée d'un domaine de connaissance et un ensemble de (SWRL) règles qui constituent ensemble un système expert.

Le support de raisonnement natif dans l'ontologie est utilisé pour déduire de nouvelle connaissance à partir des règles prédéfinies sur les transactions suspectes. Le système expert résultant a été testé avec un ensemble de 8 millions de transactions bancaires commerciales. Cette approche utilise une technique de conduite ontologique qui ne minimise pas uniquement le coût de modélisation de la connaissance mais rend le système expert extensible.

IV. CONSTRUCTION D'ONTOLOGIE

Pour construire notre ontologie, nous devons choisir une méthodologie de construction d'ontologie. Pour cela nous avons effectué une recherche bibliographique sur les méthodologies existantes. Nous avons trouvé plusieurs méthodologies de construction des ontologies tel que: Enterprise [8],Tove [9],Methontology[10] etNeOn Methodology[7]. Cette dernière a été construite suite à

lacomparaison des trois autres méthodologies (Methontology, On-To-Knowledge et Diligent).

Tove, Entreprise, Methontology et Neon Methodology sont les méthodologies les plus représentatives pour construire des ontologies. Et sont les plus utilisées. Pour construire notre ontologie nous avons utilisé la méthodologie (Neon Methodology) car elle englobe tous les scénarios possibles. Le scénario 1 : de la spécification à l'implémentation. Le scénario 2 : la réutilisation et la réingénierie des ressources non ontologiques. Le scénario 3 : la réutilisation des ressources ontologiques. Le scénario 4 : la réutilisation et la réingénierie des ressources ontologiques. Le scénario 5 : la réutilisation et la fusion des ressources ontologiques. Le scénario 6 : la réutilisation, la fusion et la réingénierie des ressources ontologiques. Le scénario 7 : la réutilisation des ontologies de modèle de conception (ODP). Le scénario 8 : la restructuration des ressources ontologiques. Le scénario 9 : la localisation des ressources ontologiques.

Nous avons implémenté notre ontologie en utilisant l'outil Protégé.

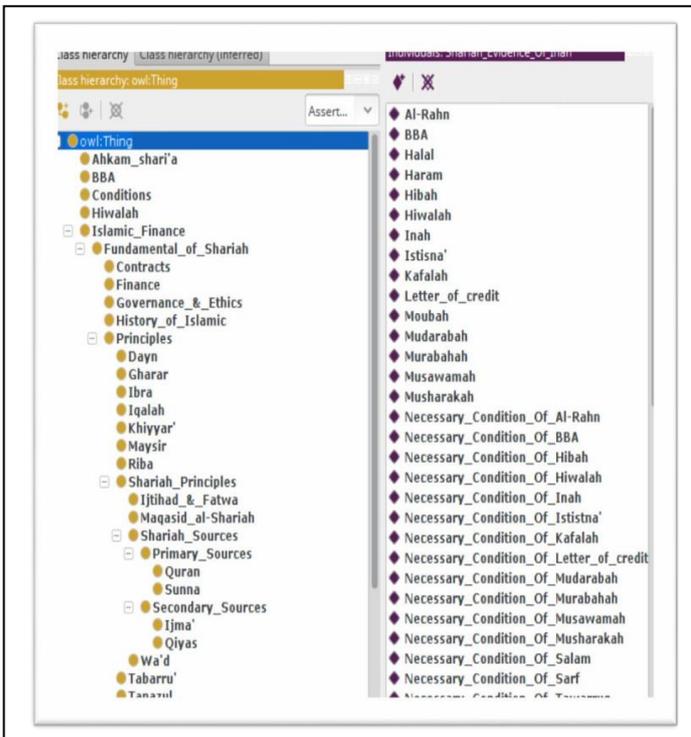


figure1. Ontologie du domaine des transactions bancaires.

La (figure 1) présente l'ontologie résultante. Cette dernière a été intégré dans notre système d'aide à la décision conçu afin d'assister El Mufti dans la génération d'El Fatwa, en adoptant l'architecture proposée dans La (figure 2). Notre système est composée d'une base des cas (où on stock tous les cas possibles), une ontologie qui contient tous les concepts du domaine avec les synonymes et un algorithme de raisonnement.

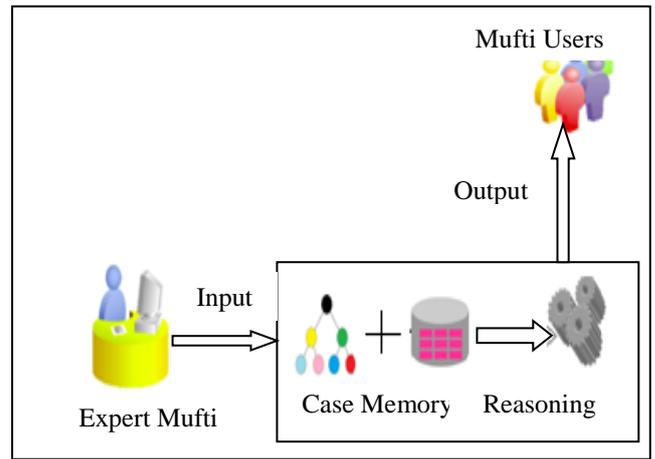


Figure2. Architecture du système d'aide à la décision basé cas

V. CONCLUSION

Nous avons conçu un système pour l'aide à la décision dans le cadre de la législation islamique. Ce système utilise le raisonnement à base des cas pour résoudre les problèmes et supporté par une ontologie de domaine. Cette dernièreest un facteur clé de succès du système. Car elle augmente son efficacité, et elle permet de représenter les connaissances du domaine d'une manière explicite indépendamment de la base de cas.

References

- [1] A. Ghazali, M. El Shafi, "El Moustasfa Fi Ilm El oussoul", ed, Dar El Koutoub El Ilmia.
- [2] N. Nouaouria, F. Atil, MT. Laskri, D. Bouyaya, AH. Amari, "A case based tool as intelligent assistant to Muft", Arabian journal for science and engineering, April 2006.
- [3] A. Amari, F. Atil, N. Bounour, N. Nouaouria, « intelligent tool for Mufti Assistance », In International Journal on Islamic Applications in Computer Science and Technology, Vol.3, Issue 2, pp: 1-9 June, 2015, ISSN 2289-4012.
- [4] A. Mamadolimova, N. Ambiah, D. Lukose, « Modeling Islamic finance Knowledge For Contract Compliance in Islamic Banking », 15th international conference, KES 2011, Kaiserslautern, Germany, September 12-13, 2011, proceeding, part 3, pages 346-355.
- [5] L. Shue, C. Chen, W. Shiue, "The development of an ontology-based expert system for corporate financial rating", Expert Systems with Application, ELSEVIER, volume 36, Issue 2, Part 1, March 2009, Pages 2130-2142.
- [6] Q. Rajput, NS. Khan, A. Larik, S. Haidar, "Ontology Based Expert-System For Suspicious Transaction Detection", Computer and Information Science, vol 7, N°1, 2014, ISSN 1913-8989, published by Canadian center of science and education.
- [7] M. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, "TheNeOnMethodology forOntology Engineering".In:Ontology EngineeringinaNetworkedWorld, pp.9--34, Springer, BerlinHeidelberg, Germany (2012)
- [8] M. Ushold, M. King, "towards a methodology for building ontologies", paper presented at the workshop on basic ontological issues in knowledge sharing, held in conduction with IJCAI-95, New York, NY.
- [9] M. Gruninger, MS. Fox, "Methodologyforthedesignandevaluationof ontologies".In: WorkshoponBasic OntologicalIssues inKnowledgeSharing, pp.6.1--6.10, Montreal, Canada (1995).
- [10] M. Fernandez-Lopez, A. Gomez-Perez, N. Juristo, "METHONTOLOGY: FromOntologicalArtTowardsOntologicalEngineering".In:AAAI SymposiumonOntological Engineering, Stanford, USA(1997)
- [11] A. Delteil, C. Faron, "A Graph-Based Knowledge Representation Language".15th European conference in Artificial Intelligence, ECAI 2002.

Analyse des sentiments appliquee aux microblogs

Youcef Benkhedda
Azouaou Faical : directeur de these
Ecole nationale Supérieure d'Informatique

Abstract—La popularite explosive des services de microblog encourage de plus en plus les internautes a partager leurs opinions et exprimer leurs sentiments en publique. Lanalyse des sentiments appliques sur ces ressources riches en opinion peut etre un moyen tres efficace pour reveler les orientation du publique et comprendre les comportement des individus. A cet effet, Loptimisation des techniques danalyse de subjectivite et des sentiments (Subjectivity and Sentiment Analysis) est une tache d'actualite qui simpose pour faire face aux defis sociaux que rencontre la societe sur les media sociaux, tel lanalyse des communautes religieuses et politiques, la lutte antiterrorisme, lamelioration du marketing virale, laide au prediction des stocks,etc.

I. INTRODUCCIÓN

Lanalyse des sentiments dans les microblogs permet de capturer les sentiments caches dans un large eventail de donnees, generale- ment de type hypertextuelle. On peut extraire lavis de laudience sur un film ou un produit recemment sortis[5],leur choix du candidat presidentiel prefere[3], leur emotion par rapport au sujet des dernieres attentats[7], etc

II. TRAVAIL REALISE ET PERSPECTIVES

Pendant la premiere annee de doctorat, nous avons effectue des experimentations sur des donnees issues du service Instagram. En pratique, nous nous sommes dedies au sujet du Halal dans le reseau Instagram en collectant tous les microblogs qui contenaient le mot Halal.Par la suite nous avons utilise des techniques danalyse statistique descriptive (probabilite,correlation, distribution,etc), des techniques de Natural language processing (Latent Dirichlet Allocation algorithm) et danalyse a base de lexicons; dans le but de determiner le sentiment des utilisateur du reseau Instagram envers le sujet Halal.

Pendant cette deuxieme annee de doctorat nous avons finis le travail fait en collaboration avec Qatar foundation sur l'analyse des donnees demographique au sein du reseau social instagram. Cela a donne resultat a un article de huit pages qui a ete soumis dans differentes conferences.

III. RESULTATS ET DISCUSSION

Le travail d'analyse de donnees avec Yelena Mejova de Qatar Foundation a ete elabore pour etre degage en un papier resumant notre travail, nous avons tente de viser des conferences tres bien classees dans le domaine du social media. A savoir la conference ICWSM17 de AIII (International Conference on Web and Social Media de Association for the Advancement of Artificial Intelligence), et WebSci '17 (International ACM Web Science Conference 2017), Malheureusement nous avons eu le rejet dans les deux conferences.

Ci-dessous un resume des reviezs des jury (metariview) issus des deux conferences.

METAREVIEW ICWSM17 Conference:

This paper studies peoples discourse of the halal culture on Instagram and finds there are qualitative differences in what halal is across the world. The reviewers find the topic to be important and agree that it needs to receive more attention. The paper is easy to follow and a pleasant read.

The reviewers make several key suggestions to improve the paper. Reviewer 2 raises an important concern about novelty and impact. The other reviewers share this sentiment in that they find the paper to lack a concrete goal and touching upon too many aspects for a single conference paper. Reviewer 1 and Reviewer 3 ask about details of the statistics and data methodology that also seem critical. The reviewers find the topic modeling description to be unclear.

Based on these reviews, I judge that the paper bares weaknesses that the authors need to address better. The topic is of great importance to the ICWSM community and the authors may try to reframe their research to have a smaller, concrete goal and a strong methodology. I hope the reviews provide helpful insights for improving the paper.

METAREVIEW WEBSCI'17 Conference :

The reviewers praised how this study is placed within the "Generation M" work and how it brings some interesting results (such as the dominance of commercial entities). However, they also noted important weaknesses suggesting in general a lack of depth in some areas. The reviewers asked for a better relationship between Halal and wellbeing/exercise/food (R1), an attempt at a better break-down by demographic group, even if demographics are only inferred (R1), and in general a richer analysis (R2). This paper can benefit from having a better balance of data and insights, which in the current form seem to be more heavy on the data. There were also some methodological questions (i.e., grounded content vs LDA topics; number of posts used) that might need some clarification but are relatively minor in comparison with the problem of lack of depth in the analysis.

Nous nous sommes ensuite enregistre dans la publication frontiers par le biais de leur topic : "Culturomics: Interdisciplinary Path Towards Quantitative Study of Human Culture". Nous avons eu quelques remarques de type "minor revise and resubmit" que nous avons vites rectifier et renvoyer au comite.

IV. CONCLUSION

Dans ce resume nous avons presente le travail effectue durant la deuxieme annee de doctorat. Nous pouvons dire que les refus de nos applications aux deux conferences ICWSM et WebSci nous ont permis d'ameliorer beaucoup d'aspects de notre travail. Cela nous a permis d'avoir une pre-acceptance dans la publication frontiers. Nous envisageons dans le future de developper notre travail afin de pouvoir publier dans des conferences mieux classees.

V. BIBLIOGRAPHY

- [1]A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media, pages 3038. Association for Computational Linguistics, 2011.
- [2]N. Godbole, M. Srinivasaiah, and S. Skiena. Large-scale sentiment analysis for news and blogs. ICWSM, 7(21):219222, 2007.
- [3]S. Hong and D. Nadler. Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience. Government Information Quarterly, 29(4):455461, 2012.
- [4]B. Liu. Sentiment analysis and subjectivity. Handbook of natural language processing, 2:627666, 2010.
- [5]A. Oghina, M. Breuss, M. Tsagkias, and M. de Rijke. Predicting imdb movie ratings using social media. In European Conference on Information Retrieval, pages 503507. Springer, 2012.
- [6]B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1135, 2008.
- [7]A. Westling. Sentiment analysis of microblog posts from a crisis event using machine learning. Master of Science Thesis, KTH CSC, Stockholm, Sweden.

Mesure de la satisfaction du décideur

Djamila Bouaissa
ENS Kouba
Kouba Alger
d_bouaissa@esi.dz
Rachid Chalal
Ecole Supérieur d'Informatique
Oued Smar Alger
r_chalal@esi.dz

Résumé

Ce présent document est une présentation de l'état d'avancement du travail de recherche qui rentre dans le cadre du doctorat portant sur la mesure de la satisfaction du décideur pour pouvoir effectuer la quatrième inscription. Nous évoquons dans ce résumé la problématique puis les travaux réalisés à ce jour et ce qui reste à faire.

I. INTRODUCTION

Le décideur est un acteur très important dans toute organisation. Pour la prise de la décision, il se base essentiellement sur l'information. Aujourd'hui dans l'aire de l'information, il est contraint par la pertinence de l'information.

Ce travail de recherche se focalise sur l'évaluation de la satisfaction du décideur quant à l'information obtenue suite à une requête, arrive-t-il à obtenir l'information qu'il faut pour le problème en question ?

La première phase de ce travail de recherche a porté sur l'exploration de l'évaluation où l'objectif était de revisiter les travaux réalisés dans l'évaluation des programmes sociaux, l'évaluation des systèmes d'information et l'évaluation des systèmes d'aide à la décision (ou décision support system ; DSS) en leur qualité de SI spécifiques qui aident ou assistent le décideur dans son processus de prise de décision. La satisfaction de l'utilisateur a été identifiée comme la mesure la plus utilisée dans la mesure du succès aussi bien des SI que des DSS.

Dans une seconde phase notre attention a porté sur l'ergonomie informatique, une science qui a pour objectif d'améliorer l'interaction homme-machine, la facilité d'utilisation et d'apprentissage des produits interactifs. Nous avons été poussé par la lacune que nous avons constaté dans l'évaluation des DSS coté interface alors que pour les logiciels les critères ergonomiques sont bien définis que cela soit par les chercheurs ou par les normes et standards. Les critères ergonomiques qui touchent à l'utilité, l'utilisabilité et la satisfaction, influent sur le succès et l'utilisation des logiciels, nous pensons que cela est aussi valable pour les DSS du fait qu'ils soient des logiciels.

Dans le présent résumé, on évoque une petite synthèse sur les concepts évoqués ci-dessus à savoir l'évaluation, l'évaluation des SI et entre autre la satisfaction de l'utilisateur, l'évaluation des DSS et l'ergonomie informatique. Nous finirons par ce qui va suivre dans le travail de recherche et une conclusion.

II. L'ÉVALUATION

Quiconque responsable de la mise en œuvre d'un projet ou d'un programme dans n'importe quel domaine est appelé un jour ou l'autre à s'interroger sur la manière dont le projet produit ses effets afin d'apporter les améliorations et les réajustements nécessaires. Cela suppose la mesure du réel afin de le juger par rapport aux objectifs fixés au départ et décider les actions à entreprendre ; c'est l'évaluation.

Notre parcours dans la littérature a permis de relever les concepts clés de l'évaluation. Elle est nait avec le souci d'évaluer des programmes éducatifs et se trouve aujourd'hui appliquée sur n'importe quel projet ou programme dans tous les domaines. Un projet ou un programme peut être évalué à n'importe quelle phase de son cycle de vie, l'objet de l'évaluation diffère d'une phase à l'autre.

L'évaluation est une démarche très rigoureuse conduite par un évaluateur pour des fins très précises. Soit pour mesurer l'impact, l'efficacité et l'efficience d'un projet ou l'aide à la décision pour des changements visés.

III. L'ÉVALUATION DES SYSTEMES D'INFORMATION

Sous les influences conjointes de la mondialisation des marchés, de la libéralisation de l'économie, de la croissance de l'activité de service et essentiellement sous l'impact des technologies de l'information et de la communication, de nouveaux modes d'organisation se sont imposés au niveau des entreprises.

Le pilotage et l'organisation de l'entreprise s'appuient sur l'information qui est au cœur du système d'information.

« Le système d'information(SI) est aujourd'hui au cœur de la création de la valeur au sein de l'entreprise. C'est un ensemble souvent hétérogène car constitué d'éléments qui se sont juxtaposés au fil du temps au gré des choix stratégiques, des évolutions technologiques des systèmes informatiques en place du développement de l'entreprise elle-même. » [5].

Pourquoi l'évaluation des SI ?

De nos jours le SI est un élément essentiel de l'entreprise. Les entreprises sont prêtes à investir de fortes sommes afin d'obtenir un SI performant et adapté à leur activité.

Cependant, comment savoir que le SI en place est adapté au à l'entreprise ? C'est l'objet de l'évaluation des SI. Elle permet de vérifier la contribution de SI à la performance de l'entreprise et favorise le progrès dans l'organisation.

Dans la littérature « évaluer un SI » est aussi abordée sous « mesurer le succès d'un SI ».

L'évaluation du SI porte essentiellement sur les variables permettant de mesurer le succès suite à une implantation (évaluation ex-post). Une faible importance est accordée aux

facteurs critiques de succès étudiés lors de la phase conception du SI (évaluation ex-ante).

DeLone et McLean ont élaboré un modèle d'évaluation devenu aujourd'hui un standard. Ils ont proposé six dimensions d'évaluation du succès du SI. La réussite du SI est fonction de sa qualité générale, la qualité de l'information, la qualité du service, son utilisation, la satisfaction des utilisateurs et les bénéfices nets. Ces dimensions sont interdépendantes ; la qualité du système ainsi que la qualité de l'information affectent de façon individuelle et collective l'utilisation ainsi que la satisfaction des utilisateurs.

La satisfaction de l'utilisateur

La satisfaction de l'utilisateur est la mesure la plus utilisée dans l'évaluation du succès des SI à cause essentiellement des impacts intangibles des SI sur une organisation qu'on ne peut pas mesurer avec des mesures objectives. C'est un concept inspiré du domaine de la psychologie et introduit dans l'évaluation du succès des SI dès les premières années d'introduction des technologies de l'information, une période qui a connu l'échec de mise en place des SI et où aussi les managers essayaient de prouver et justifier les investissements en SI. C'est un concept qui a été largement évoqué et critiqué dans la littérature portant sur l'évaluation des SI mais d'une façon « one point at one time ». En vue de faciliter sa compréhension, nous avons élaboré une synthèse schématique sur ce concept que nous avons mis en évidence dans un modèle de classe du formalisme UML. Ce travail a fait l'objet d'un article qui a été accepté et présenté dans la conférence CCIT 2017 à Zurich.

Nous avons aussi implémenté dans le cadre d'un PFE (2017) le modèle de mesure de la satisfaction de l'utilisateur de Bailey et Pearson tel que proposé par les auteurs en 1981. Le logiciel développé a été exploité dans le cadre d'un mémoire de Master (2017) pour mesurer la satisfaction des utilisateurs du SI opérationnel de la sécurité sociale.

IV. L'EVALUATION DES DSS

DSS is the field of IS discipline that is focused on supporting and improving managerial decision making. DSS is about developing and deploying IT based systems to support decision

processes. The aim of early DSS developers was to create an environment in which the human decision maker and the IT-based system worked together in an interactive fashion to solve problems; the human dealing with the complex unstructured parts of the problem and the information system providing assistance by automating the structured elements of the decision situation. The purpose of the development of DSS is an attempt to improve the effectiveness of the decision maker.

Evaluation is the process that measures the success of DSS. The measurement of a DSS success can be accomplished through the assessment of four domains; effectiveness, efficiency, use and user satisfaction. Maynard, Bustein and Arnott have identified a set of criteria useful for measuring the success of DSS projects within decision-maker and management. They presented it as a hierarchy classified by the different perspectives from which DSS is measured (effectiveness, efficiency, use and user satisfaction). According to this classification and the state of the art about DSS success, we have classified the criteria dealing with DSS user satisfaction in seven dimensions: Information quality, DSS quality, decision-maker, organizational services, Decision process, decision task and Interface (Design). Our class model about user satisfaction in IS research is under adaptation to DSS user satisfaction.

V. PERSPECTIVES

L'exploration de la satisfaction de l'utilisateur nous a permis d'élaborer une vue multidimensionnelle de la satisfaction de l'utilisateur dans la recherche en SI. Notre objectif est d'adapter cette nouvelle vision pour la mesure de la satisfaction du décideur en considérant les dimensions identifiées dans la mesure de la satisfaction des DSS selon leurs caractéristiques et leurs spécificités par rapport aux SI.

VI. CONCLUSION

Dans ce résumé, nous avons présenté d'une façon très brève les travaux abordés dans le cadre de notre travail de recherche sur la mesure de la satisfaction du décideur.

Robustesse de la reconnaissance d'activité

Maamar Melaini

Directeurs de thèse : M. Yacine Challal

Équipe : (LMCS)

Résumé— Dans le processus de la reconnaissance d'activité à base de deep learning qui se train pour classifier les activités humains, on utilise une méthode métaheuristique, l'algorithme des abeilles (ABC) pour avoir une solution optimimale des poids et rapide.

I. INTRODUCTION

Dans l'algorithme des abeilles, il y a trois type d'abeilles : les butineuses actives, inactives et les éclaireuses, l'emplacement de la source de nourriture représente la solution possible au problème, et la quantité du nectar de cette source correspond à une valeur objective dite **fitness**.

Dans l'article [1] l'algorithme hybride de colonie d'abeilles artificielles (ABC) et Multi-couche Perceptron (MLP) ont été proposés pour construire un IDS de réseau efficace. Le MLP a été utilisé comme classificateur pour distinguer les paquets normaux et anormaux du trafic réseau. La structure de MLP a été créée en s'appuyant sur les caractéristiques du jeu de données (NSL-KDD 99). De plus, l'algorithme ABC est utilisé pour la formation de MLP en optimisant les valeurs de pondération et de biais.

Dans l'article [2] l'auteur utilise l'algorithme ABC qui a de bonnes capacités d'exploration et d'exploitation dans la recherche d'un jeu de poids optimal afin de former le réseau de neurones. Car les algorithmes de formation traditionnels présentent certains inconvénients tels que la réduction des minimums locaux et la complexité informatique.

[3] Dans cet article, l'auteur présente un modèle de MNN basé sur l'algorithme ABC (ABC-MNN). Une comparaison entre ABC-NN et ABC-MNN montre que le modèle ABC-MNN a une plus grande précision de formation et une performance de généralisation.

[4] dans cet article l'auteur présente une méthode qui traite données biologiques. La récupération des données nécessite une application systématique des méthodes expérimentales et de calcul. Les résultats comparatifs produits par les systèmes S et les ABC-NN, GA-NN, PSO-NN ont été étudiés. Les résultats de la simulation montrent que ABC-NN proposée surpassait GA et PSO en terme de précision et leurs résultat converge vers les système S.

[5] Cet article présente une variation de l'algorithme ABC original, à savoir la coopérative Article Bee Colony (CABC), qui améliore significativement l'ABC d'origine pour résoudre des problèmes complexes d'optimisation. L'algorithme CABC est utilisé pour optimiser six fonctions de référence largement utilisées et les résultats comparatifs produits par ABC, Particle Swarm Optimization (PSO) et sa version coopérative (CPSO) ont été étudiés. Les résultats de la simulation ont montré que la CABC proposée surpassait les trois autres algorithmes en termes de précision, de robustesse et de vitesse de convergence.

[6] Dans cet article, les auteurs présentent une méthodologie basée sur ABC, qui maximise la précision et minimise le nombre de connexions d'une ANN en développant en même temps les poids synaptiques, l'architecture de l'ANN et les fonctions de transfert de chaque neurone. La méthodologie est testée avec plusieurs problèmes de reconnaissance de formes.

[7] dans cet article un algorithme hybride combinant l'algorithme de colonie d'abeilles artificielles (ABC) avec l'algorithme de Levenberg-Marquardt (LM) est introduit pour former des réseaux de neurones artificiels (ANN). La formation d'une ANN est une tâche d'optimisation où l'objectif est de trouver un ensemble de poids optimal du réseau dans le processus de formation. Les algorithmes de formation traditionnels pourraient être bloqués dans les minimums locaux et les techniques de recherche globales pourraient réduire considérablement les minima globaux. Par conséquent, des modèles hybrides combinant des algorithmes de recherche globale et des techniques conventionnelles sont utilisés pour former des réseaux de neurones.

Dans l'article [8] La combinaison de l'algorithme ABC et du BP du réseau de neurone est adoptée de sorte qu'un modèle non linéaire peut être identifié et qu'un pendule inversé puisse être contrôlé. Les résultats de la simulation montrent que la capacité de cartographie étendue du réseau de neurones et l'algorithme de convergence global rapide d'ABC peuvent être obtenues en combinant l'algorithme de colonie d'abeilles artificielles et le réseau de neurones

Dans l'article [9] présente un réseau de neurones MLP qui nécessite généralement une grande quantité de mesures disponibles afin d'obtenir une bonne précision de classification. Pour remédier à cela, les auteurs utilisent un réseau de neurones de liaison fonctionnelle (FLNN) qui possède une seule couche de poids de connexion entraînaibles. La méthode standard pour régler le poids dans FLNN utilise un algorithme Backpropagation (BP-learning). Pourtant, l'algorithme d'apprentissage BP a des difficultés telles que le piégeage dans l'optique locale et la convergence lente qui peuvent affecter les performances de FLNN. Dans ce travail, un algorithme de colonie d'abeilles artificielles (ABC) connu pour avoir de bonnes capacités d'exploration et d'exploitation dans la recherche de poids optimal est utilisé pour récupérer les inconvénients de l'apprentissage BP. Avec les modifications apportées aux comportements de forgeresse de l'abeille occupée et des spectateurs, la mise en œuvre de l'ABC modifiée comme un schéma d'apprentissage pour FLNN a permis un meilleur taux d'exactitude pour résoudre les tâches de classification

II ALGORITHME ABC

Les butineuses sont attribués aux différentes sources de nourriture de façon à maximiser l'apport total de nectar. La

colonie doit **optimiser** l'efficacité globale de la collecte. La répartition des abeilles est donc en fonction de nombreux facteurs tels que la quantité du nectar et la distance entre la source de nourriture et la ruche.

Le nombre des butineuses actives ou inactives représente le nombre de solution dans cette populations.

Dans la première étape, l'algorithme génère une population initiale de SN solutions distribuées de façon aléatoire. Chaque solution x_i ($i = 1, 2, \dots, SN$) qui est initialisée par les éclaireuses, et représente un vecteur de solution au problème d'optimisation. Les variables que contient chaque vecteur doivent être optimisées.

Après l'initialisation, la population des solutions est soumise à des cycles répétés $C = 1, 2, \dots, C_{max}$, ces cycles représentent des processus de recherches faits par les butineuses actives, inactives et les éclaireuses.

Les butineuses actives recherchent dans le voisinage de la source précédente x_i de nouvelles sources v_{ij} ayant plus de nectar, Elles calculent ensuite leur fitness. Afin de produire une nouvelle source de nourriture à partir de l'ancienne, on utilise l'expression ci contre :

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj})$$

Où $k \in \{1, 2, \dots, BN\}$ BN est le nombre des butineuses actives et $j \in \{1, 2, \dots, SN\}$ sont des indices choisis au hasard. Bien que k est déterminé aléatoirement, il doit être différent de i . ϕ_{ij} est un nombre aléatoire appartenant à l'intervalle $[-1, 1]$, il contrôle la production d'une source de nourriture dans le voisinage de x_{ij} .

Après la découverte de chaque nouvelle source de nourriture v_{ij} , un mécanisme de sélection gourmande est adopté, c'est-à-dire que cette source est évaluée par les abeilles artificielles, sa performance est comparée à celle de x_{ij} . si le nectar de cette source est égale ou meilleur que celui de la source précédente, celle-ci est remplacée par la nouvelle. Dans le cas contraire l'ancienne est conservée.

Pour un problème de minimisation, La fitness est calculée suivant cette formule :

$$fit_i(\bar{x}_i) = \begin{cases} 1 & \text{si } f_i(\bar{x}_i) \geq 0 \\ 1 + f_i(\bar{x}_i) & \text{si } f_i(\bar{x}_i) < 0 \end{cases}$$

Telle que $f_i(\bar{x}_i)$ est la valeur de la fonction objectif de la solution \bar{x}_i .

A ce stade, les butineuses inactives et les éclaireuses qui sont entrain d'attendre au sein de la ruche. A la fin du processus de recherche, les butineuses actives partagent les informations sur le nectar des sources de nourriture ainsi que leurs localisations avec les autres abeilles via la danse frétillante. Ces dernières évaluent ces informations tirées de toutes les butineuses actives, et choisissent les sources de nourriture en fonction de la valeur de probabilité P_i associée à cette source, et calculée par la formule suivante :

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n}$$

Où fit_i est la fitness de la solution i , qui est proportionnelle à la quantité du nectar de la source de nourriture de la position i .

La source de nourriture dont le nectar est abandonné par les abeilles, les éclaireuses la remplacent par une nouvelle source. Si durant un nombre de cycle prédéterminé appelé « limite » une position ne peut être améliorée, alors cette source de nourriture est supposée être abandonnée.

Toutes ces étapes sont résumées dans l'algorithme :

Entrée : S, W, O

Sortie : la meilleure solution

1. Initialiser la population avec S+W solutions aléatoires
2. Evaluer la fitness de la population
3. **Tant que** le critère d'arrêt n'est pas satisfait **faire**
 4. Recruter O butineuses inactives et attribuer chacune à un membre de la population
 5. **Pour** chaque butineuse inactive affectée à un membre de la population **faire**
 - a. Effectuer une itération de l'algorithme de recherche de nouvelle source
 6. Evaluer la fitness de la population
 7. **Si** un membre de la population ne s'est pas amélioré au cours des itérations **faire**
 - a. Sauver la solution et remplacer la par une solution aléatoire
 8. Trouver S solutions aléatoires et remplacer les S membres de la population qui ont la mauvaise fitness
9. **Fin Tant que**
10. **Retourner** la meilleure solution

Algorithme de l'optimisation par colonie d'abeille :

S : nombre de butineuse éclaireuse.

W : nombre de butineuse active.

O : nombre de butineuse inactive.

II. TRAVAUX REALISE

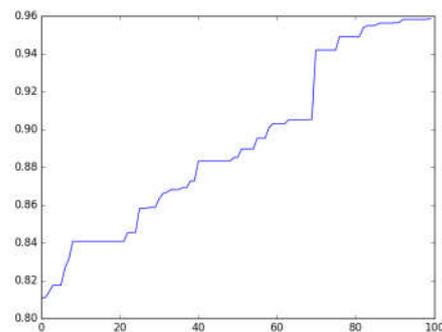


Figure 1 : La solution initiale est aléatoire

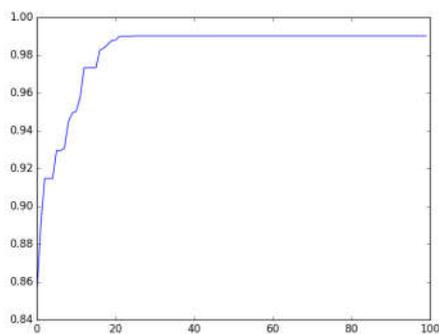


Figure 2 : La solution initiale est aléatoire plus l'ajustement avec la solution optimale de chaque itération

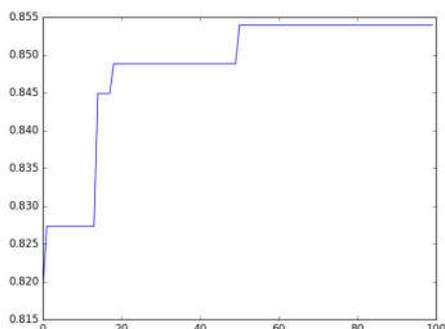


Figure 3 : La solution initiale est optimale plus l'ajustement avec le voisinage de la solution optimale de chaque itération

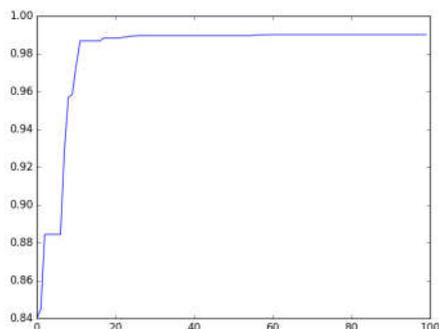


Figure 4 : La solution initiale est optimale plus l'ajustement avec la solution optimale de chaque itération

III. CONCLUSION

Références

- [1] Hybrid Intrusion Detection System Using Artificial Bee Colony Algorithm and Multi-Layer Perceptron. MAHMOD S. MAHMOD 2015
- [2] Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks. Dervis Karaboga 2007
- [3] Artificial Bee Colony Algorithm for Modular Neural Network. Chen Zhuo-Ming 2013
- [4] Artificial bee colony algorithm-neural networks for S-system models of biochemical networks approximation. Wei-Chang Yeh 2010
- [5] Cooperative Approaches to Artificial Bee Colony Algorithm. Wenping Zou 2010
- [6] Artificial Neural Network Synthesis by means of Artificial Bee Colony (ABC) Algorithm Beatriz A. Garro 2011
- [7] Hybrid Artificial Bee Colony Algorithm for Neural Network Training. Celal Ozturk 2011
- [8] Back Propagation Neural Network Based on Artificial Bee Colony Algorithm. Feihu Jin, Guang Shu 2012
- [9] Solving a Classification Task using Functional Link Neural Networks with Modified Artificial Bee Colony Yana Mazwin Mohamad Hassim 2013

Directeur de thèse
M. Yacine Challal

Doctorant : Lotfi Bouziani

Laboratoire : LMCS

Sujet de thèse :

Optimisations des techniques analytiques, dans un contexte Big Data.

Résumé :

Devant l'accroissement des volumes de données collectées et les besoins de traitements de plus en plus consommateurs/gourmands en ressources et en temps, l'utilisation des techniques traditionnelles d'accès et de consultation de données s'avère insuffisante.

En effet, dans ce contexte Big Data, les données sont généralement non-structurées, hétérogènes et à granularité variable, nécessitant une réadaptation des techniques et des méthodes liées à leur traitement et à leur exploitation. Parmi les défis majeurs à relever, on citera l'intégration et la persistance des données et des modèles, la visualisation des flux de données en temps réel et l'analyse des données d'une manière efficace et évolutive.

Pour répondre à ces besoins, MapReduce a été proposé comme un paradigme évolutif, capable de traiter de grands volumes de données. Il est considéré comme l'une des approches les plus prometteuses pour répondre aux besoins d'analyse et de traitement, permettant une exécution parallèle et distribuée sur un grand nombre de nœuds. Il a déjà été implémenté dans plusieurs projets Open Source (comme dans Apache Hadoop).

Cependant, MapReduce rencontre aussi plusieurs limitations, comme par exemple, l'absence d'un langage de requêtes de haut niveau tel que SQL, le type de stockage distribué à utiliser pour les données (RDBMS ou NoSQL), la complexité générée par les analyses interactives et les besoins itératifs dans l'apprentissage automatique (machine Learning), le traitement des flux de données en ligne (Data Stream processing), la sécurité des données et le respect de la vie privée lors des traitements.

Dans cette thèse, nous nous proposons d'étudier et d'étendre le paradigme MapReduce, pour relever certaines des limitations relatives aux analyses (Big Data Analytics). Il s'agira de proposer des améliorations des techniques existantes et des solutions pour répondre aux limitations rencontrées. L'optimisation sera étudiée en fonction d'un certain nombre de critères, tel que le temps d'exécution, la qualité et la quantité de données échangées (cas des traitements distribués) et l'énergie consommée.

Mots clés :

Big Data, MapReduce, Analyses et requêtes, NoSQL, Hadoop.

Avis du Directeur de thèse

Avis de co-Directeur de thèse

L'Apprentissage des Agents dans un Environnement Orienté Education

Mohamed Redha Sidoumou

Directeurs de thèse : Prof Kamal Bechkoum, Prof Karima Benatchba

Équipe : (CoDesign, Optimisation, MSI, EIAH, Image, SURES)

Résumé—In this thesis the emotion and behaviour of drivers are simulated. The aim is to reduce Road rage via simulation. For that we focus on the emotional aspect of driving rather than the abilities of individual drivers. To simulate this phenomenon, emotion models are studied and modelled computationally. Then, the influencing factors are collected from different studies. Attention theories are used to simulate the multitasking nature of drivers. A machine learning technique is proposed to select the behaviour.

I. INTRODUCTION

Emotion simulation is a field of work that lies between Artificial Intelligence and Psychology. Its goal is to attempt to reproduce the same interactions that happen in the organism during an emotional episode. This kind of simulation has many outcomes and possibilities. It can enable virtual characters to go from preprogrammed synthetic emotions to dynamically generated emotions for more realism. This can be used in virtual learning environment and games. It can also allow testing psychological theories where evaluation is a sensitive part. Another potential application is to be used in robotics.

To create an implementable model an overview of the systems involved in the emotional process should be dressed. For this, studying different models explaining the emotion phenomenon is required. Then, the next step is converting these models into concrete models by exploring the abstract parts of their abstract parts using assumptions and sometimes other theories.

We can find different models in the literature. Some are derived from psychology, others from neuroscience and physiology. In our study, we focus mainly on psychological models since they provide a link between the external world and the mind. These settings are the closest to allow a concrete implementation of an emotional system.

One of the obstacles we face when simulating emotions is selecting an appropriate model. Models can differ and sometimes contradict each other. However, looking at them from different perspectives can allow reconciliation between few models and allow a combination of models to achieve the required simulation.

The selected models need to be converted into implementable models. Machine learning techniques are used for such an implementation.

I. CONTEXT

Current computational emotional models use a simple paradigm that from one situation derives one single emotion and then from both of these entries triggers a coping strategy which is an action to be executed by the individual. This scenario does not reflect the multitasking nature of the human body and mind. Humans can multitask [1]. Furthermore, an emotional episode can take several steps before a stability is reached. This is shown on in [2] where an event can trigger a series of emotional

modifications instead of only one. Because of these gaps, we propose a model that enables multitasking and progressive emotional change for the sake of realism in simulating emotions. After the completion of the model a machine learning algorithm is used to determine appropriate actions to be triggered. The used machine learning algorithm is presented below.

II. COMPLETED WORK

A. Using attention theories for multitasking

Early attention theories depict it as a single pool of resources like in (Kahneman, 1973) where it can be shared among different tasks. In those theories, the resources are similar and each task requires a certain amount of resources from that pool. More recent theories suggest that humans have multiple types of resources and allocate those resources according to the attended tasks. In other words, a person can attend successfully to two tasks at the same time if those two tasks use different types of resources or if they require the same resource and the amount required by each task can be satisfied at the same time. One of those theories is Wicken's, which is described in detail in (Wickens, 2005).

An algorithm based on Wicken's theory has been implemented to determine whether two or more tasks can be executed at the same time thus enabling a policy for allowing multitasking when dealing with multiple events or executing them.

B. Implementing Scherer's theory

Scherer's theory describes the emotional phenomenon as a process. This process involves multiple changes in the affective state instead of a discreet change. However, this theory is defined at a higher abstraction level and is not ready to be implemented as it is. For that, multiple theories like the PDA (Pleasure dominance, Arousal) [3] for affective state representation. Although this theory competes with the appraisal theories which encompasses Scherer's theory, a common ground was possible bridging the gap between the two schools. Several other theories were used as well to make Scherer's model more concrete.

C. A Machine learning technique: Classification via Projection

This machine learning technique try to find the optimal separating hyperplane. The found hyperplane is similar the one found by the SVM. The idea is that instead of looking for a separating hyperplane, we look for a hyperplane where the projection of the points in the spaces is perpendicular to the optimal separating hyperplane. With this idea, we just need to optimize over one variable. P is what we are looking for, p_a and p_b are the projection of the points belonging to class A and class

B respectively over (P). (A) is the separating hyperplane that is perpendicular to (P). What we need is to get the largest possible distance between pa and pb.

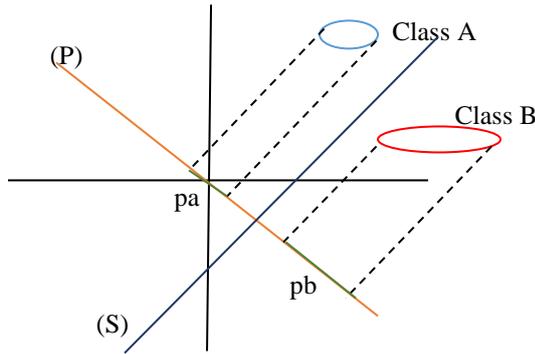


Figure 1 Projection example

To find the optimized (P) we proceed we need the following formulas:

The projection of the points is calculated by:

$$P(x, y) = x \cdot \cos(\alpha) + y \cdot \sin(\alpha)$$

Where α is the angle between (OX) and (P)

Therefore the function to optimise is:

$$MAX_{\alpha}(MIN_{i,j}(x_i - x_j) \cos(\alpha) + (y_i - y_j) \sin(\alpha))$$

Where $i \in class A indexes$ and $j \in class B indexes$

Example :

In a two-dimensional space, pt1(1,1), pt2(0,0), pt3(2,0), pt4(1,-1) are points. pt1 and pt2 belongs to the class A. pt3 and pt4 belong to the class B. To find the optimal separating hyperplane we should maximise the following formula:

$$MIN \begin{cases} (1 - 2) \cos(\alpha) + (1 - 0) \sin(\alpha) \\ (1 - 1) \cos(\alpha) + (1 + 1) \sin(\alpha) \\ (0 - 2) \cos(\alpha) + (0 - 0) \sin(\alpha) \\ (0 - 1) \cos(\alpha) + (0 + 1) \sin(\alpha) \end{cases}$$

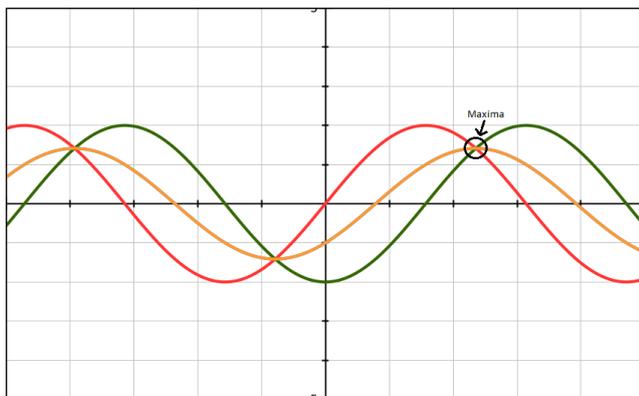


Figure 2 Functions plotting

The problem with the function above is the MIN which is not derivable. However, the function $a \cdot \cos(x) + b \cdot \sin(x)$ is convex when ≥ 0 . The minimum of convex functions is always convex thus a variation of a gradient descent applied to all of the functions will give the required optimization.

This technique works well when the two groups of points are linearly separable. This technique also proved efficient to predict whether two groups of points are linearly separable or not.

To predict if two groups of points are linearly separable or not we need only to look if the function to optimize can have a positive value by studying its sign.

This technique has been used to select actions by drivers during an emotional episode.

D. Evaluation

An evaluation of the implemented model has been carried out. The model was connected to a synthetic face to display facial characteristics of emotions. 19 Students evaluated the displayed emotions in various scenarios. The evaluators answered questions related to the gender of the represented drivers and the realism of the simulation.

Results showed that using the progressive emotional modifications can account for individual differences such feminine traits.

III. CONCLUSION

The emotional process has still many secrets to disclose. Theorists are far from reaching any agreements. These setting make it quite challenging to implement. The used model is a combination of models that may compete against each other. However, different perspective in implementing them can lead to reconcile some aspects.

Implementing Scherer's theory is challenging since it is linked to the cognitive process. The model need to determine at what speed the appraisal valued are determined by cognition. Theories addressing this aspect could not be found.

Also, using personality in the model has been challenging. Personality variables were included in the model. However, results showed that these variables did not have much differences in the believability of the simulation.

An interesting finding is that progressive emotional change was able to show some personality traits like femininity/masculinity.

REFERENCES

- [1] C. D. Wickens, "Multiple resources and performance prediction," *Theor. Issues Ergon. Sci.*, vol. 3, no. 2, pp. 159–177, Jan. 2002.
- [2] K. Scherer, "Appraisal considered as a process of multilevel sequential checking," in *Appraisal processes in emotion: Theory, methods, research*, K. R. Scherer, A. Schorr, and T. Johnstone, Eds. New York: NY: Oxford University Press, 2001, pp. 92–120.
- [3] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*. Cambridge, Mass: The MIT Press, 1974.

Assistance des apprenants dans un jeu sérieux Basée sur l'analyse de traces

Dahmani Rabea

Directeurs de thèse : Faical Azouaou, Bousbian Nabila

Équipe : EIAH

Résumé— Les Jeux Sérieux constituent aujourd'hui un outil pédagogique utilisé aussi bien dans les parcours d'apprentissage initial que dans les formations en entreprise ou continues. La question de l'efficacité de ce type de dispositifs d'apprentissage constitue un enjeu de recherche. L'une des questions de recherche les plus importantes est relative à l'analyse du comportement de l'apprenant joueur à partir de ses traces d'interaction. L'objectif de cette analyse est d'évaluer son apprentissage et l'assister en cas de difficulté. Nous cherchons à proposer un outil de suivi et de remédiation pour un apprenant dans un dispositif d'apprentissage hybride basé sur les jeux sérieux.

Introduction.

Julian Alvarez [10] définit un jeu sérieux comme étant une « *Application informatique, dont l'objectif est de combiner à la fois des aspects sérieux (Serious) tels, de manière non exhaustive, l'enseignement, l'apprentissage, la communication, ou encore l'information, avec des ressorts ludiques issus du jeu vidéo (Game). Cette association doit s'opérer par l'implémentation d'un scénario pédagogique.* ». L'objectif de l'utilisation des jeux sérieux est de favoriser la motivation et l'engagement de l'apprenant joueur [1]. Dans ce cas la question de l'évaluation des apprentissages par le jeu, s'impose. Qu'apprend-on précisément en jouant ? [3]. Pour évaluer cet apprentissage faut-il alors suivre l'apprenant, en traçant ses actions[4].

Notre objectif principal est de proposer un dispositif permettant d'intégrer les jeux sérieux dans un scénario d'apprentissage, et d'utiliser les traces afin d'améliorer l'apprentissage et faciliter la remédiation pour les apprenants joueurs.

Problématique

Notre travail traite des problématiques suivantes :

1. Comment intégrer un jeu sérieux dans un dispositif d'enseignement universitaire ? comment favoriser l'apprentissage ? quoi proposer ? à quel moment ? et comment ?
2. Quelle méthodologie utiliser pour passer d'un contenu pédagogique vers un contenu ludique ? comment faire la modélisation des connaissances ?
3. Comment créer un jeu capable de s'adapter au profil du joueur/apprenant et aux évolutions des connaissances enseignées ?

4. Quels sont les principaux ressorts du jeu garantissant la motivation du joueur ?
5. Comment évaluer et assister l'apprenant joueur ?

. Travaux réalisés

La première année de thèse a été consacrée à:

1. L'élaboration d'un scénario d'apprentissage hybride intégrant un jeu sérieux, fondé essentiellement sur des séquences de cours et d'exercices (Dispositif d'enseignement hybride).
2. L'élaboration d'un scénario de jeu « brut » en électronique fondamentale. modélisation des aspects pédagogiques et ludique ; Il s'agit de passer d'un référentiel de compétences à des éléments scénarisés intégrant la question de la jouabilité (le gameplay).
3. Recherche bibliographique concernant [5][6][7][8][9]:
 - La Conception des jeux sérieux.
 - Les outils d'analyse de traces dans les jeux sérieux.

Pour cette deuxième année :

1. Nos premières réflexions sur les différentes méthodologies permettant d'intégrer un apprentissage dans un jeu, à savoir le « game design » lié aux aspects ludiques et la modélisation des connaissances ; nous ont permis la gamification d'un parcours d'apprentissage ainsi que la réalisation d'un prototype de mini jeu en électronique fondamentale.
2. Définition d'une nouvelle approche pour la conception du scénario de jeu. Ce qui permet un meilleur respect du rythme d'apprentissage. Apprentissage interactif, compatibilité avec apprentissage en ligne et/ou à distance. Apprentissage semi-autonome.
3. Conception d'un modèle (global) d'assistance pédagogique du jeu, basé sur l'évaluation de l'apprenant joueur.

- Perspectives

Notre objectif pour l'année prochaine :

- Améliorer le jeu. Résoudre le problème de progression. Quels problèmes devront être résolus pour avancer dans le jeu? Par quel coefficient de difficulté soutenir la motivation des joueurs? Il importe ici d'évaluer le niveau de connaissances des étudiants avant le jeu, mais aussi celui que l'on veut qu'ils atteignent.
- définir un système d'assistance actif et personnalisé, capable d'identifier les apprenants dans des situations difficiles et de fournir une assistance appropriée.

Conclusion

Pour apporter une réponse à notre problématique, nous avons exploré les approches existantes dans la littérature concernant la conception et l'intégration des jeux sérieux dans des dispositifs d'apprentissage. Notre approche de solution est une première ébauche, qui tente de définir un modèle de conception capable de répondre à nos besoins.

[2] Egenfeldt-Nielsen S. « Practical barriers in using educational computer games ». *On the Horizon*. 1 mars 2004. Vol. 12, n°1, p. 18- 21.

[3] Connolly T., Stansfield M., Hainey T., Cousins I., « Arguing for Multilingual Motivation in Web 2.0: A Game-Based Learning Platform for Language Learning ». In : *European Conference on Games Based Learning*, Graz, Autriche, 2009.

[4] Settouti L., Prié Y., Mille A., Marty J. C., « Systèmes à base de traces pour l'apprentissage humain ». In : *TICE*, Toulouse, France , 2006.

[5] Marfisi-Schottman I., George S., Frank T.-B. « Tools and Methods for Efficiently Designing Serious Games ». In : *Proceedings of ECGBL 2010 The 4th European Conference on Games Based Learning. 4th ECGBL*. Danish School of Education Aarhus University, Copenhagen, Denmark, 2010. p. 226- 234.

[6] Marfisi-Schottman I. *Méthodologie, modèles et outils pour la conception de Learning Games*. Thèse de Doctorat en Informatique. Lyon, France : Université Claude Bernard Lyon 1, 2012. 339 p.

[7] Mariais C. *Modèles pour la conception de Learning Role-Playing Games en formation professionnelle*. Thèse de Doctorat en Informatique. Université de Grenoble, 2012.

[8] Marne B., Labat J.-M. « Implémentation de patrons de conception pour adaptation des parcours pédago-ludiques dans les jeux sérieux ». In : *Actes du 8ème Colloque Technologies de l'Information et de la Communication pour l'Enseignement. TICE 2012*. Lyon, France, 2012. p. 69- 79.

[9] Marty J., Carron T., Pernelle P. « Observe and react: interactive indicators for monitoring pedagogical sessions ». *International Journal of Learning Technology*. 1 janvier 2012. Vol. 7, n°3, p. 277 - 296.

10

[1] Prensky M. « Digital game-based learning ». Octobre 2003, vol. 1, n°1, 2003.

Un accès contextuel adaptatif et proactif aux systèmes d'aide de décision mobile: approche basée sur l'utilisation d'un modèle flexible d'un profil contextuel sémantique.

Ramzi Lemouchi

Directeur de thèse : Admane Lotfi

Équipe : LMCS (ESI)

Résumé—la logistique est élément décisif et important pour mener n'importe quel type des opérations militaires. Le temps est très important pour prendre des décisions critique d'approvisionnement et de soutien logistique selon la situation en cour des forces d'armées est en temps réels. Cela nécessite le développement d'un système d'aide à la décision logistique mobile (Mobile Logistic Decision Support System –MLDSS-)

I. INTRODUCTION

La disponibilité des technologies de communication et de l'information a donné une large utilisation des systèmes d'aide à la décision. De plus, le développement des connexions sans fils et des dispositifs mobiles peut complètement changer la façon de prendre la décision pas seulement n'importe où et n'importe quand mais surtout dans les situations ou le temps de prendre la décision est critique. Par exemple les DSS pour cas médical d'urgence ou le temps de prendre la décision est très critique et doit éviter toute sortes d'erreurs, ou DSS pour prendre la décision comment agir face à un feu de forêt selon les conditions météorologiques en temps réel pour éviter la propagation du feu sur des régions voisins sensibles.

Dans les systèmes décisionnels à temps critique le changement instantané des données d'entrées actuelles impacte les résultats (les données de sorties) ou la relation entre ces changements est une caractéristique importante pour prendre une décision. Ainsi, les décideurs ont besoin des informations à la fois robuste à la situation actuelle et à jour en terme de temps et position instantanées et à évaluer les risques de prendre une décision dans ce moment (la connexion est possible) ou dans un moment ultérieur (peut être la connexion sera perdue). Par exemple si un problème non structuré (stratégique) le temps pourrait être moins critique que dans le cas d'un problème semi-structuré (Opérationnelle) donc plus il y a du temps pour raffiner et améliorer la qualité des données avant que la décision finale soit prise.

Le but principale de les systèmes d'aide à la décision mobile MDSS (Mobile Décision Support System) est de permettre aux utilisateurs de gérer facilement la base de données ou de connaissance à travers des connexions sans fils en utilisant des différents dispositifs mobiles (les téléphones intelligents, les PDAs, les ordinateurs portables, les tablettes,...etc.). Donc, on peut dire que le but principal d'un système décisionnel mobile à temps critique est de fournir des informations (des indicateurs) convenables et utiles dans des délais très important pour assister les utilisateurs (les professionnelles) à prendre des décisions correctes et rentables.

II. EVALUATION DES MDSS

Les MDSS sont entrain de propager dans différents domaine d'application (médicale, commerciale, sécuritaire, militaire,...etc.). Par contre, certaines lacunes empêchent la croissance et la mise en œuvre de ces systèmes dans l'environnement réel. Ces lacunes sont des sujets de différents travaux de recherche pour les combler. Par exemple *Padmanabhan et al* [1] ont souligné certains obstacles devant la croissance et l'intégration des MDSS, tels que la confidentialité, l'interopérabilité, la capacité de stockage et de traitement des dispositifs mobiles et le manque de standards internationaux et des Framework d'évaluation.

L'évaluation des MDSS est un problème qui persiste depuis l'existence des DSS classique car il y peu de travaux de recherche autour de ce sujet et il n'y avait pas de modèle d'évaluation bien défini de ce type de systèmes. *Emna et al* [2] ont proposé un ensemble de critères essentiels pour l'évaluation d'un MDSS en soulignant ceux qui sont liés au contexte de l'utilisateur. Ces critères sont regroupés en deux catégorie selon deux facteurs de qualités (1) l'utilité ou l'efficacité qui est un facteur lié à l'aspect fonctionnel des MDSS et (2) l'intérêt ou l'importance qui est lié à l'évaluation de la qualité de la connaissance extraite et son adéquation avec le contexte d'utilisation.

Les fonctionnalités des MDSS peuvent être vérifiées à travers les cinq (05) caractéristiques qui sont définies dans la norme ISO 9126 comme suit [3] :

L'Interopérabilité : c'est un critère clé, particulièrement dans le cas de la mobilité, qui signifie le nombre d'occurrence de manque de communication influant l'échange de données et de commandes entre ce système et d'autres systèmes ou équipements qui sont interconnectés.

La Sécurité : c'est le critère le plus important dans le cas des systèmes mobile et militaires puisque les connexions sont sans fils et les données sont très sensibles. Il signifie le nombre d'occurrence des problèmes de sécurité durant le test ou l'utilisation, par exemple la perte des informations sensible, les accès et les opérations illégal et les fuites de sécurité des donnés de sorties.

La Précision : la fréquence d'occurrence des résultats incorrects ou imprécis dû aux insuffisances de données et l'incohérence entre les fonctions et leur description dans le manuel d'utilisateur.

La Pertinence : représente le niveau d'occurrence d'opérations ou de fonctions insatisfaisante (ne fournit pas le résultat attendu de l'utilisateur pour accomplir sa tâche principale) durant le la phase de test ou et l'utilisation du système.

La Conformité : mesure le nombre d'occurrence des fonctions qui ne respectent pas les normes, les conventions, les contrats et d'autres exigences règlementaires. C'est un critère qui affecte l'utilité du système.

De plus la qualité de la connaissance extraite par un MDSS peut être évaluée à travers six (06) critère principale selon le model proposés par *Emna et al* [2] puisqu'il n'existe pas auparavant un modèle ou une norme pour d'évaluation des MDSS. Ces critères sont décrits comme suit :

- ▣ **L'Assistance** : le niveau de complétude et d'exactitude avec lequel l'utilisateur atteint ces objectives spécifiés.
- ▣ **La Complexité** : le niveau de complexité de l'algorithme d'exploration des données (Data Mining) qui décrit le temps nécessaire pour l'exécution de cet algorithme. Ce critère est très important dans les systèmes utilisés dans les situations critiques (sous pression du temps).
- ▣ **La Résistance aux Bruits** : permet de vérifier s'il y a des données d'entrées raté ce qui implique les connaissances extraites et les données de sortie.
- ▣ **La Nouveauté** : permet de vérifier si la connaissance extraite n'est pas déjà connu par l'utilisateur.
- ▣ **La Compréhensibilité** : décrit la capacité d'interprétation de la connexion extraite par l'utilisateur.
- ▣ **La connaissance préalable** : permet de vérifier si la connaissance extraite nécessite d'autres connaissances additionnelles pour que l'utilisateur puisse l'interpréter.

Chaque critère d'évaluation représente un aspect de recherche pour améliorer le fonctionnement et l'utilité des MDSS. Par contre, dans le cas des MDSS militaires y as trois critères principales qui doivent être fortement assurés sont la sécurité, la complexité et la nouveauté. Par exemple, pour détruire une rompe de lancement de missile ennemi à l'aide d'un MDSS il est nécessaire d'avoir l'information correcte dans le temps réel et dans le niveau de détail nécessaire. Donc, la tâche de décision dans ce cas est compliquée, l'utilisateur a besoin de décider comment cibler et quand détruire la cible ennemi et utiliser quel type d'arme tout en assurant la confidentialité de la source de données et de son dispositif.

Dans le domaine militaire la logistique est un aspect très important influant sur le résultat de n'importe quelle opération et pour n'importe quelle formation militaire d'armée. La logistique militaire est l'ensemble des actions qui visent à soutenir les opérations des forces armées. Pour l'Organisation du Traité de l'Atlantique Nord (OTAN), la logistique est la science de la planification et de l'exécution de différentes opérations des forces armées et de leur maintenance. Généralement, la logistique recouvre les taches suivants :

- ▣ conception, expérimentation, acquisition, maintenance et réparation des matériels et des équipements ;
- ▣ transport du personnel, des matériels et des équipements ;
- ▣ acquisition, construction et entretien d'installations et d'infrastructures ;
- ▣ ravitaillement en combustibles, en vivres et en munitions ;
- ▣ acquisition ou prestation de services ;
- ▣ soutien médical et sanitaire.

C'est bien pendant la seconde guerre mondiale que la logistique prend toute son envergure. En effet, lors du débarquement en Normandie, il a fallu non seulement débarquer 150 000 hommes sur les plages mais aussi près d'une quinzaine de milliers d'hommes embarqués par avions ou planeurs, et ce, en seulement quelques heures. Le jour même a nécessité une logistique parfaite, toute erreur étant fatale. D'après le Général Dwight David Eisenhower «La logistique influence toutes les batailles – elle est même souvent décisive». A ce propos, le général De Gaulle déclara : « C'est en respectant la logistique que le général Eisenhower mena jusqu'à la victoire ».

Aujourd'hui, toutes les armées sont conscientes de l'importance d'un appui logistique. La grande mobilité, la technicité, la force de frappe et la rapidité des forces d'armées les obligèrent à constamment perfectionner la logistique militaire. Dans ce but, les institutions militaires de différentes nations ont constitué un pôle de réflexion et d'action dans le domaine de la gestion des flux des matériels et d'articles d'approvisionnements (acquisition, stockage et attribution), tout en insistant plus particulièrement d'instaurer des réglementations et des approches pour la maîtrise optimisé et le bon contrôle de ces flux. Cela a abouti à développer un système de codification et d'identification des matériels qui met en évidence un langage d'approvisionnement unique entre les différents échelons de soutien logistique.

Dans notre travail nous essayons de concevoir un Système D'aide à la Décision Logistique Mobile (Mobile Logistic Decision support system –MLDSS-) tout en respectant les trois (03) critères d'évaluation principaux cités précédemment.

References

- [1] N. Padmanabhan, F. Burstein, L. Churilov, J. Wassertheil, B.Hornblower, and N. Parker, A Mobile Emergency Triage DecisionSupport System Evaluation, In IEEE Proc. of the 39th Hawaii International Conf. on System Sciences, Kauai, January 2006p.
- [2] Emna et al, Context aware criteria for the evaluation. IEEE ICIS, Las Vegas, USA, 2015.
- [3] Software Product Evaluation-Quality and Guidelines for Their Use, Geneva International Organisation for Standards Characteristics, 2002

Vers un Simulateur de Conception Physique de Bases de Données Sémantiques

Nabila Berkani

Directeurs de thèse : Ladjel Bellatreche

Équipe de Recherche :

Résumé— La forte volumétrie des données décrites par des ontologies a conduit à la naissance des bases de données sémantiques (BDS). Les communautés industrielles et académiques se sont intéressées de près à cette technologie, où plusieurs solutions ont été proposées pour représenter, stocker les données sémantiques au sein des SGBD et manipuler ses données via des langages de requêtes comme Sparql. Parallèlement, le cycle de vie de conception des bases de données et entrepôts de données a convergé vers les étapes suivantes : analyse des besoins, phase conceptuelle, phase logique, ETL et la phase physique. La mise en place d'un simulateur nécessite la proposition d'une approche de conception qui suit ce cycle. Notre travail consiste en la proposition d'une approche de conception des bases de données (entrepôts de données) revisitant l'ensemble des phases du cycle de vie permettant un déploiement à la carte. Dans cette optique, un simulateur sera mis en place pour gérer les différents types de bases de données sémantiques.

I. INTRODUCTION ET PROBLEMATIQUE

Les travaux de notre thèse de doctorat s'inscrivent dans le domaine des bases de données Sémantiques (BDS) et s'intéressent plus particulièrement à la problématique générale de la conception.

On assiste aujourd'hui à une augmentation importante des volumes de données à traiter par les différentes applications d'entreprise. Le développement des réseaux, et notamment d'internet, a rendu omniprésent le besoin de manipuler et d'analyser des données provenant de diverses sources réparties, autonomes et hétérogènes. Les bases de données s'avèrent une technologie adaptée à une telle activité. Cependant, les concepteurs les construisaient pour répondre à un besoin particulier, sans se soucier si ces bases de données seraient intégrées et utilisées par des personnes en dehors de leur entreprise. Résultat, elles se heurtent au problème d'hétérogénéité structurelle et sémantique des données. Les ontologies ont été utilisées pour résoudre ce problème en explicitant le sens des concepts et des relations.

En effet, les ontologies sont considérées comme des modèles conceptuels qui permettent de représenter explicitement la sémantique des données. Elles peuvent être considérées comme une solution appropriée pour faire face au défi principal de construction des bases de données, à savoir l'hétérogénéité. Les ontologies sont aujourd'hui utilisées dans un nombre croissant d'applications. En conséquence, des quantités importantes de données sémantiques appelées aussi données ontologiques sont produites quotidiennement. Initialement, ces données sémantiques étaient gérées par des outils en mémoire centrale. Avec la multiplication des ontologies de domaine et le volume important de données manipulées, est apparu le besoin de persister ces données ontologiques dans des bases de données, appelées bases de

données à base ontologiques (BDBO). Différentes BDBO ont été proposées : RDFSuite, Sesame, Jena, Oracle, OntoDB, ..., OntoMS. Elles diffèrent principalement dans l'architecture et le modèle de stockage utilisé. En réalité, l'utilisation croissante des ontologies a donné lieu à la proposition de différents formalismes (RDF, OWL, PLIB), où chacun cible un domaine particulier, ce qui a engendré une diversité des modèles de stockages utilisés. On distingue trois représentations] : verticale, binaire et horizontale. La représentation verticale stocke les données dans une table unique à trois colonnes (sujet, prédicat, objet). La représentation binaire stocke les classes et les propriétés dans des tables différentes. La représentation horizontale traduit chaque classe comme une table ayant une colonne pour chaque propriété de la classe.

Devant cette diversité des modèles de stockage, différentes architectures ont été proposées. D'abord les architectures de type I qui sont similaires à celle des bases de données traditionnelles distinguant deux parties : partie méta-schéma et partie données où la structure de l'ontologie (concepts et propriétés) et les instances ontologiques sont stockées. Les architectures de type II séparent le modèle d'ontologie de ses données distinguant ainsi trois parties : le méta-schéma, le modèle d'ontologie et le schéma de données. Puis sont apparues les architectures de type III offrant plus de flexibilité à la partie ontologie en ajoutant la partie méta-modèle réflexive de l'ontologie. Cette évolution des architectures et des modèles de stockages utilisés fait des BDBOs de réelles candidates pour les projets de déploiement.

Par ailleurs, ces dernières années la communauté de recherche a convergé vers un cycle de vie de construction des bases de données à savoir : (i) définition des besoins, (ii) modélisation conceptuelle, (iii) modélisation logique, (iv) modélisation physique et (v) déploiement. Notre travail s'inscrit dans cette optique où nous avons constaté qu'il existe très peu de travaux ayant proposé une approche de conception traitant de l'ensemble des étapes du cycle de vie et permettant un déploiement à la Carte selon des différents types de BDBO. La proposition d'un simulateur est bien appropriée pour mieux maîtriser leur diversité.

Par ailleurs, si de nombreux travaux ont été menés sur la conception physique dans le contexte des bases de données traditionnelles, peu se sont intéressés à la conception physique dans les BDS. Après une analyse approfondie des principales BDS existantes, nous affirmons que leur conception physique est plus complexe. Cela est dû à leur diversité qui porte sur :

(1) des formalismes supportés: chaque BDS utilise un formalisme particulier pour définir ses ontologies (OWL, PLIB ou FLIGHT); (2) des modèles de stockage: contrairement aux bases de données traditionnelles, où le modèle logique est stocké selon une approche relationnelle,

dans une BDS, une variété de modèles de stockage (représentation horizontale, spécifique, verticale, etc.) sont utilisés pour stocker des ontologies et des instances ontologiques et (3) des architectures des SGBD cible supportant ces bases. Trois types d'architecture existent : (a) architecture "deux quarts" des BD traditionnelles. Dans cette architecture les ontologies et les données sont stockées ensemble. (b) Le second type d'architecture qualifié de "trois quarts" sépare les ontologies des instances ontologiques ; (c) la dernière architecture dite "quatre quarts" ajoute à la précédente une partie appelée méta-schéma. Pour mieux maîtriser cette diversité, et proposer une approche de conception à la carte, le développement d'un simulateur est bien approprié.

II. ETAT DE L'ART

Dans cette partie, nous présentons l'état de l'art étudié concernant la problématique traitée.

A. Etat de l'art sur les BDBO

Dans cette section, nous allons présenter de manière sommaire les bases de données à base ontologique en abordant les architectures et modèles de stockages permettant la représentation des ontologies au sein d'une base de données.

Définition d'une BDBO

Une Base de Données à Base Ontologique (BDBO) [16] est une base de données qui contient des ontologies, des données et des liens entre ces données. Une BDBO possède les caractéristiques suivantes : (1) l'ontologie et les données sont représentées dans une même base de données et peuvent faire l'objet des mêmes traitements (insertion, mise à jour...) ; (2) toute donnée est associée à un élément ontologique qui en définit le sens et inversement ; et (3) l'ontologie locale à la base de données possède éventuellement des références vers des ontologies externes.

L'architecture 1 d'une BDBO qui possède les parties qu'on retrouve dans les bases de données classiques à savoir : le catalogue système et les données stockées, s'ajoutant à cela une ontologie locale qui peut référencer une ontologie externe.

Ces BDBO permettent de stocker les données au niveau ontologie (classes et propriétés) qui contient la sémantique des données et les données au niveau instance appelées données à base ontologiques (DBO) [Fankam 2009]. Une DBO est une donnée qui référence une classe d'une ontologie qui en définit la sémantique. Les références aux classes et propriétés sont définies par des identifiants universels dont dispose chaque modèle d'ontologie.

Ces dernières années, différentes architectures de BDBO ont été proposées : RDFSuite [2], Sesame [4], Oracle [6], OntoDB [15, 6], OntoMS [15]. L'ensemble de ces architectures différent dans :

- Le modèle d'ontologies supporté,
- Le schéma de base de données utilisé pour stocker les ontologies,

- Le schéma de base de données utilisé pour représenter les données à base ontologique,
- Les mécanismes de définition des liens entre les données (instances) et les éléments des ontologies (classes, propriétés).

Architectures de représentation des ontologies dans les BDBO

Selon la représentation des ontologies dans les BDBO [8], trois architectures ont été proposées :

- BDBO de type I

Dans ce type d'architecture, un seul schéma de représentation, décrit dans la figure 5.2, est utilisé pour stocker les classes, propriétés et instances de l'ontologie. Le modèle de stockage est vertical. Les concepts (classes et propriétés) de l'ontologie et les DBO (instances) sont décrits suivant le formalisme RDF en utilisant la structure (sujet, prédicat, objet).

L'approche est basée sur l'utilisation d'une table unique à trois colonnes (sujet, prédicat, objet), nommée table verticale [1]. L'insertion dans la table de triples est une opération simple. Cependant, l'accès aux instances d'une classe et aux valeurs des propriétés qui les caractérisent nécessite de nombreuses auto-jointures sur la table.

- BDBO de type II

Dans ce type d'architecture, les DBO et l'ontologie sont stockées dans deux schémas de représentation différents, décrits dans la figure 5.5. Cette approche est basée sur la définition d'une table pour chacun des concepts (classes, propriétés) et axiomes (subsumption de classes) selon le modèle d'ontologie supporté.

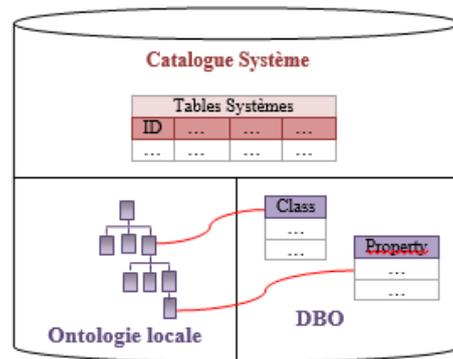


Fig. 1. : Architecture de la BDBO de type II.

Cette représentation est dépendante du modèle d'ontologies supporté par la BDBO. Elle est qualifiée de spécifique. Elle n'autorise pas l'introduction de concepts provenant d'autres modèles d'ontologies. L'approche est statique et non évolutive. Ce type d'approche de représentation spécifique est adopté par les BDBO : (1) Sesame, RDFSuite et RStar basées sur le modèle RDF-Schema ; (2) OntoMS, DLDB et PARKA basées sur OWL et (3) KAON basée modèle KAON.

- BDBO de type 3

Dans ce type d'architecture, les DBO et l'ontologie sont stockés dans deux schémas de représentation différents de la même façon que dans les BDBO de type II, rajoutant à cela un autre schéma appelé méta-schéma, décrit dans la figure 5.7. La présence du méta-schéma offre une flexibilité à l'ontologie qui est représentée comme une instance du méta-schéma.

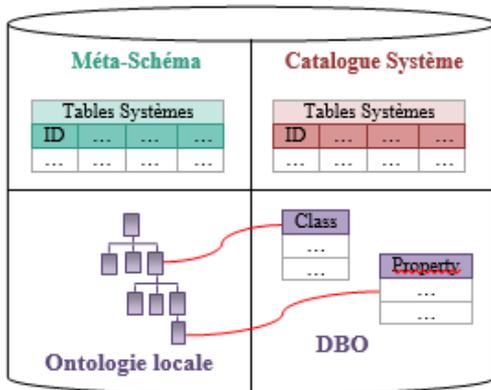


Fig. 2.: Architecture de la BDBO de type II.

L'avantage de l'ajout d'une partie méta-schéma est de permettre : [9]

- Un accès générique aux ontologies,
- Une évolution du formalisme d'ontologie utilisé,
- Un stockage de différents formalismes d'ontologies (OWL, DAML+OIL, PLIB...).

B. Etat de l'art sur la phase ETL

Nous avons étudié les différents travaux de conception ETL.

Dans [21], les auteurs ont construit une ontologie pour automatiser le processus ETL. Ils ont utilisé un raisonneur pour identifier les opérations nécessaires pour charger des données. Les auteurs dans [20] proposent un mécanisme de langage naturel à base de modèles pour décrire les sources de données et les opérations ETL dans un rapport narratif textuel plus approprié pour l'utilisateur. Les deux œuvres supposent que le schéma des sources de données et entrepôt de données sont déjà connus et ils ne traitent que de phase ETL.

Les auteurs dans [18] ont défini un schéma de conception multidimensionnelle à partir d'une ontologie OWL décrivant les sources de données. Ils ont identifié le concept de fait (relatif à l'attribut mesure) et les concepts de dimensions qui y sont reliés par des relations one-to-many. La sortie est un schéma en étoile ou flocon de neige apte à être instancié dans une base de données multidimensionnelle traditionnelle. Les auteurs dans [12], ont proposé une méthode formelle basée sur l'utilisation d'OWL/DL pour l'identification des concepts multidimensionnelle (fait et dimension). L'analyste définit le modèle multidimensionnel de l'ontologie de domaine existant.

Romero et al. ont décrit dans [17] une méthode qui combine la modélisation conceptuelle multidimensionnel et la conception ETL. Ils commencent par une analyse des sources

de données mises en correspondance à une ontologie de domaine. Ensuite, ils définissent les étapes qui mènent à l'identification des faits, des dimensions et des hiérarchies de dimension. La validation de l'utilisateur des concepts proposés est nécessaire avant de produire un schéma multidimensionnel et un flux d'opérations ETL conceptuels. Cependant les phases logiques et physiques de conception ne sont pas prises en compte dans ce travail.

[13] considère les données sémantiques fournies par le Web sémantique et annotés par ontologies OWL, à partir de laquelle un modèle DW est défini et peuplé. Cependant, le processus ETL dans ce travail dépend d'un format d'instance spécifique (triplets). [3] propose une approche de conception basée sur l'ontologie, où ontologies sont confrontés à chaque phase du cycle de vie.

La principale limite des approches discutées est le fait qu'ils utilisent une seule ontologie avec une taille réduite et peu de restrictions. Cependant, les applications réelles actuelles, référencent différentes ontologies de domaine avec une grande taille considérable au niveau d'un entrepôt de données déployé sur des BDBO.

C. Etat de l'art sur la conception physique des BDBO

Dans cette section, nous allons présenter de manière sommaire les travaux réalisés sur la conception physique des bases de données sémantiques, notamment les modèles de coûts proposés ainsi que les structures d'optimisation étudiées.

Dans [10], les auteurs proposent des estimations de la cardinalité des patrons de graphes RDF à l'aide de statistiques en utilisant le schéma de l'ontologie. Tous les sous-graphes de l'ensemble de données peuvent être déduits à partir du graphe du schéma de l'ontologie et annotés avec leurs cardinalités exactes dans la base de données.

Dans [18], les auteurs proposent une stratégie d'estimation de cardinalité construite sur des statistiques basées sur des prédicats. Les auteurs utilisent une fonction de probabilité pour estimer la cardinalité du résultat d'une requête sur un ensemble de données d'une ontologie.

Dans [22], les auteurs proposent un modèle de coût pour optimiser les patrons de graphe RDF. Ce modèle utilise la sélectivité des patrons de triplets. La sélectivité d'un patron de triplets est estimée en multipliant la sélectivité de chacun de ses composants (sujet, prédicat et objet). Le calcul des statistiques est fonction de la taille de l'ensemble de données.

Dans [14], les auteurs ont élaboré une stratégie pour construire des statistiques pour tous les sous-graphes possible d'un ensemble de données RDF dans un schéma d'ontologie. Ils se sont intéressés aux sous-graphes de formes arbitraires (graphes en forme de chaîne, d'étoiles, etc.). Ils utilisent la technique d'indexation afin d'améliorer le traitement des requêtes. Leur approche repose en grande partie sur les capacités d'indexation natives.

Dans [11], les auteurs ont proposé une approche de conception physique en suivant les étapes suivantes :

- Etude expérimentale des BDBO afin d'identifier les différents facteurs qui peuvent influencer la conception physique et surtout les caractéristiques de chaque type de base de données. Les auteurs ont utilisé six bases de données sémantiques, trois académiques (Jena, Sesame et OntoDB) et trois

issues de milieu industriel, à savoir Oracle, DB2RDF et IBM SOR.

- Développement d'un modèle de coût pour chaque type de BDBO.
- Considération d'une structure d'optimisation à savoir les vues matérialisées. Ce choix a été justifié comme suit : (i) elle a été largement étudiée dans le cas des entrepôts de données, (ii) une vue peut être indexée et partitionnée, et (iii) elle a été étudiée récemment dans le cas des bases de données sémantiques.

III. TRAVAUX REALISES

Dans un premier temps, notre travail s'est porté sur la proposition d'une approche de conception des entrepôts de données, déployés sur une BDBO. Les sources de données sont représentées par des bases de données sémantiques qui référencent une base de connaissance [4]. Le travail a été publié dans la 18^{ème} conférence internationale sur le Big Data Analytics and Knowledge Discovery - DaWaK 2016 (Classe B). L'architecture de la solution proposée est décrite dans la figure 4.

En effet, le Big Data a largement contribué à accélérer le développement des bases de connaissances volumineuses et précieuses (KB) par les académiciens (par exemple, Cyc, DBpedia, Freebase et Yago) et industriels (par exemple, Knowledge Graph). D'autre part, des études sérieuses ont identifié le rôle crucial de KB pour les tâches d'analyse, en offrant des détails sur plusieurs entités (personnes, lieux, produits, etc.). La disponibilité de ces grande KB peut contribuer à la conception d'approches à valeur ajoutée pour les applications de business intelligence. Dans cette optique, nous avons proposé une nouvelle approche pour la conception d'entrepôt de données sémantique qui considère KB dans le cycle de vie. Deuxièmement, sur la base de la formalisation proposée des KB, nous produisons la conception multidimensionnelle conceptuelle et un processus ETL sémantique orchestrant les données des flux provenant de sources de données vers le stockage DW. Enfin, toutes les étapes de notre approche sont illustrées à l'aide du YAGO KB et déployés dans Oracle graphe RDF sémantique 12c.

La solution est basée sur l'utilisation de la structure graphe permettant la gestion d'un volume important de données. Comme étude de cas, nous avons utilisé le système de gestion des bases de données Oracle 12c qui permet la gestion des données RDF en utilisant le langage d'interrogation des données ontologique Sparql.

Les principales étapes de l'approche sont :

- Définition des besoins utilisateurs. Nous avons exploité le modèle proposé au niveau du laboratoire LIAS [24].
- Définition du concept *contexte* permettant l'extraction d'une partie de la base de connaissance. Cette extraction se fait sur la base des besoins utilisateurs définis précédemment. Un algorithme a été mis en place pour extraire la partie des KB.
- Etape de modélisation conceptuelle. Une approche a été mise en place pour identifier les concepts multidimensionnelle et dériver le schéma conceptuel

MD de l'entrepôt de données sur la base de la KB et besoin des utilisateurs. Un algorithme a été mis en place pour l'automatisation de cette étape.

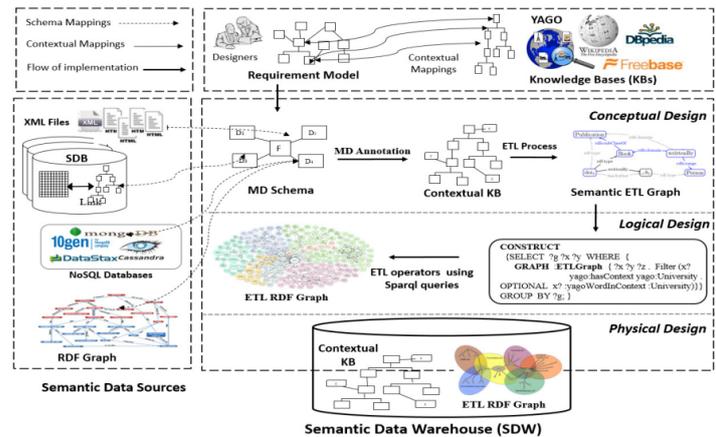


Fig. 3.: Architecture de solution proposée.

- Une étape de conception ETL a été proposée en utilisant la structure de graphe. Un algorithme ETL prend en compte en entrée les sources de données, la KB contextuelle et le schéma de l'entrepôt de données et permet de générer un graphe ETL qui représente le contenu de l'entrepôt de données (schéma + instances).
- Les étapes logique et physique ont été traitées brièvement pour produire le schéma physique.
- Une étape de déploiement a été mise en place. Un déploiement à la Carte pour permettre l'utilisation de la base de données sémantique choisie par le concepteur selon le besoin.
- Une étude de cas utilisant Oracle 12c a été proposée. Ce choix est justifié par le fait que Oracle offre les features RDF Semantic Graph et Oracle Spatial.
- Une étude expérimentale a été réalisée. Nous avons utilisée la KB Yago. À l'heure actuelle, YAGO compte plus de 10 millions d'entités (comme les personnes, les organisations, les villes, etc.) et contient plus de 120 millions de faits sur ces entités. Elle a atteint une précision de 95%. Compatible avec RDF, elle permet la génération de données dans différents formats : Turtle, N-Triple, Literals, etc. YAGO contient des connaissances sur le monde réel, y compris les entités (par exemple, l'université, les gens, villes, pays, etc.) et des faits sur ces entités (ville de chaque université, programme de chaque niveau dans l'université, quelle ville se trouve dans un pays, etc.).

Une étude expérimentale a été réalisée pour montrer l'apport de notre approche. Une comparaison avec le travail réalisé en Magister a montré l'intérêt d'utiliser la structure graphe et les bases de connaissances. Dans ce travail, les deux phases de modélisation conceptuelle, logique et ETL ont été traitées en détail.

Un second travail a été réalisé sur la phase analyse des besoins [23]. Il a été publié dans la conférence ISO/IEC 7^{ème} conférence internationale sur leveraging applications of formal methods, verification and validation). L'architecture suivante décrit les étapes de création d'un entrepôt de besoins.

Dans ce travail nous avons montré l'intérêt de créer un entrepôt de données orienté besoin. En effet, dans l'élaboration des entrepôts de données, les premières études se sont basées uniquement sur les données, et ils ont négligé les besoins des utilisateurs (UR) qui représentent une étape dans le processus de développement, qui a été la cause d'un taux d'échec important dans la conception des entrepôts de données. Cette phase souffre de nombreux inconvénients tels que : vocabulaires et formalismes hétérogènes, relations complexes entre les exigences. Cependant, les travaux des exigences d'ingénieur (RE) se concentrent sur le cycle des exigences de la vie telle que l'intégration et l'analyse, mais il n'existe pas de travail ayant essayé de traiter les exigences de la même manière que les données avec un processus d'extraction, transformation et chargement (ETL). À cette fin, ce travail propose une solution ETL dédiée aux exigences sémantiques, ce qui donne l'occasion de soutenir les exigences de la même manière que les données de niveau conceptuel.

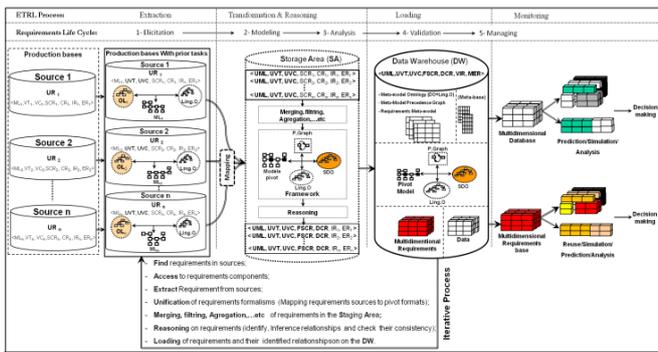


Fig. 4.: Etapes de création de l'entrepôt de besoins.

Les principales étapes de ce travail sont :

- Définition du schéma multidimensionnel en se basant sur les besoins pour dériver les concepts : mesure, fait, dimension, hiérarchie de dimension.
- Formalisation du processus ETL adapté au contexte de besoin sur la base du modèle de besoin proposé au niveau du laboratoire LIAS.
- Définition de la signature des opérateurs ETL sur la base du modèle de besoin.
- Mise en place d'un algorithme ETL besoin exploitant les opérateurs ETL définis précédemment.
- Création d'un cube de besoin en utilisant le vocabulaire QB4OLAP (proposition de la W3C).
- Déploiement de l'entrepôt de besoin sur la base de données sémantique de Oracle 12c.
- Etude expérimentale proposée pour montrer l'intérêt de la solution.

La mise en place de ces deux travaux a été validée par un outil. La solution de déploiement à la Carte sur les bases de données sémantiques est décrite en figure 7.

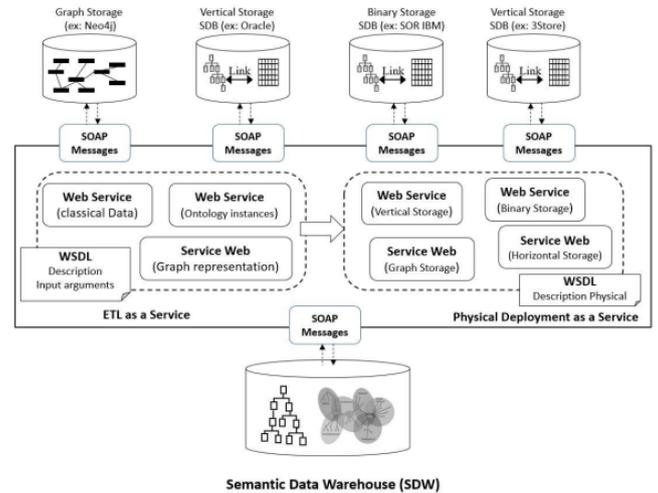
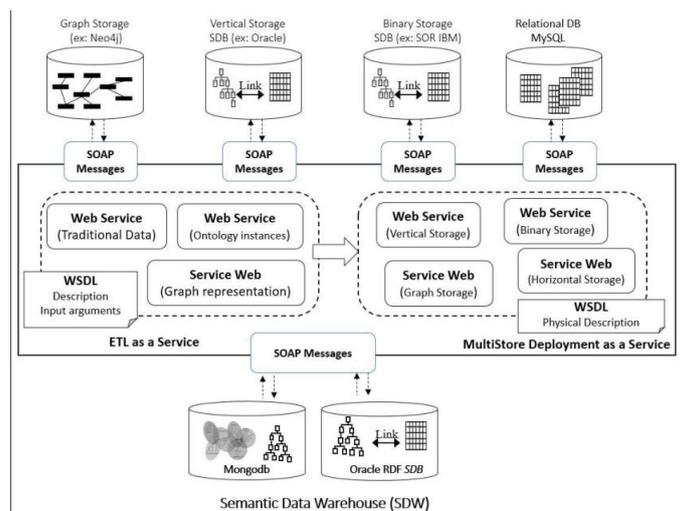


Fig. 5.: Architecture de déploiement sur les bases de données sémantiques.

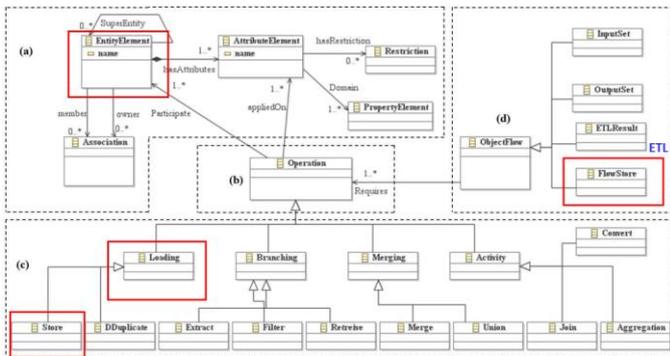
Un troisième travail a été réalisé sur la partie ETL. Ce travail a été publié dans la conférence DEXA 2017 (28th International Conference on Database and Expert Systems Applications).

Ce travail permet de réaliser la surcharge des transformations ETL. De nos jours, les petites, moyennes et grandes entreprises sont fascinées par les projets de Big data pour répondre à leurs besoins d'analyse. Dans ces projets, l'intégration de données joue un rôle crucial dans la construction de systèmes d'intégration de données. L'intégration des données doit faire face à la diversité des sources apportées par le phénomène Big Data.



Pour ajouter de la valeur, plusieurs sources de données sont construites à partir des bases de connaissances (KB) tels que YAGO. Pour intégrer les données provenant de ces sources, un processus ETL est généralement utilisé et mettant en œuvre un ensemble d'opérateurs assez bien définis qui sont sensibles aux types de sources. Cette situation oblige les développeurs à coder un processus ETL de chaque type de source.

Dans le contexte de la variété des sources de données, ce scénario doit être interrogé pour permettre à chaque opérateur ETL d'avoir plusieurs définitions, chacune ayant un type de signature différente. Ceci peut être réalisé par une surcharge des opérateurs ETL - des fonctionnalités définies dans le contexte des langages de programmation. Dans ce travail, nous avons d'abord surchargé les opérateurs canoniques de l'ETL. Deuxièmement, une approche web service mise en place pour orchestrer le flux ETL considérant différents types de sources de données. En troisième lieu, une procédure de fusion qui permet de fusionner les différentes instances des sources de données hétérogènes est présentée. Enfin, une étude de cas et un outil sont mis en place en utilisant le benchmark LUBM, la base de connaissance YAGO KB et déployé sur Oracle 12C pour les graphes RDF sémantique.



IV. PERSPECTIVES

Comme perspective à notre travail, nous sommes entrain d'effectuer une étude sur la phase physique afin de proposer une approche d'optimisation. Cette dernière permettra d'appliquer une structure d'optimisation sur les bases de données sémantiques.

V. REFERENCES

[1] R. Agrawal, A. Somani, Y. Xu. Storage and Querying of E-Commerce Data. In Proceedings of the 27th International Conference on Very Large Data Bases (VLDB), pages 149–158. Morgan Kaufmann Publishers Inc, 2001.

[2] S. Alexaki, V. Christophides, G. Karvounarakis, D. Plexousakis, K. Tolle. The ICS-FORTH RDFSuite: Managing Voluminous RDF Description Bases. In Proceedings of the 2nd International Workshop on the Semantic Web, pages 1–13, 2001.

[3] N. Berkani, L. Bellatreche, and S. Khouri. Towards a conceptualization of ETL and physical storage of semantic data warehouses as a service. Cluster Computing, 16(4):915–931, 2013.

[4] N. Berkani, L. Bellatreche, B. Benatallah: A Value-Added Approach to Design BI Applications. DaWaK 2016: 361-375, 2016.

[5] Broekstra, A. Kampman, F. van Harmelen. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In Proceedings of the 1st International Semantic Web Conference (ISWC),

number 2342 in Lecture Notes in Computer Science, pages 54–68. Springer Verlag, 2002.

[6] E. I Chong, S. Das, G. Eadon, J. Dtrinivasan. An efficient sql-based rdf querying scheme. In VLDB. Proceedings of the 31st international conference on Very Large data bases, pages 1216-1227. VLDB Endowment, 2005.

[7] H. Dehainsala, G. Pierra, L. Bellatreche. OntoDB: An Ontology-Based Database for Data Intensive Applications. In Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA), volume 4443 of Lecture Notes in Computer Science, pages 497–508, Springer, 2007.

[8] C. Fankam, L. Bellatreche, H. Dehainsala, Y. Ait Ameer, G. Pierra. SISRO: Conception de bases de données à partir d'ontologies de domaine. A apparaître dans Technique et science informatiques (TSI), pp. 1- 29, 2008.

[9] C. Fankam. OntoDB2 : un système flexible et efficace de Base de Données à Base Ontologique pour le Web sémantique et les données techniques, Thèse de Doctorat, Université de Poitiers, 2009.

[10] A. Maduko, K. Anyanwu, A. P. Sheth, and P. Schliekelman, "Estimating the cardinality of rdf graph patterns," in WWW (C. L. Williamson, M. E. Zurko, P. F. Patel-Schneider, and P. J. Shenoy, eds.), pp. 1233–1234, ACM, 2007.

[11] Bery Mbaïoussoum, Ladjel Bellatreche, Stéphane Jean, Materialized View Selection Considering the Diversity of Semantic Web Databases, Proceedings of the 18th ADBIS, pp. 163-176, LNCS, Springer, Ohrid, Republic of Macedonia September 2014,

[12] V. Nebot, R. B. Llavori, J. M. P'erez-Mart'inez, M. J. Aramburu, and T. B. Pedersen. Multidimensional integrated ontologies: A framework for designing semantic data warehouses. J. Data Semantics, 13:1–36, 2009.

[13] V. Nebot and R. B. Llavori. Building data warehouses with semantic web data. Decision Support Systems, 52(4):853–868, 2012.

[14] G. Neumann, Thomas et Weikum, "Rdf-3x: a risc-style engine for rdf," Proc. VLDB Endow., vol. 1, pp. 647–659, Aug. 2008.

[15] M. Park, J. Lee, J. Lin, O. Serres, O. Chung. An Efficient and Scalable Management of Ontology. In Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA), volume 4443 of Lecture Notes in Computer Science, pages 975–980. Springer, 2007.

[16] G.Pierra, Y. Ait-Ameer, L. Bellatreche, H. Dehainsala, S. Jean, C. Fankam, D. N. Xuan. Données à base ontologique: gestion, interrogation, intégration, Première édition des Journées Francophones sur les Ontologies (JFO 2007), Octobre, 2007

[17] O. Romero, A. Simitsis, and A. Abell'ó. GEM: requirement-driven generation of ETL and multidimensional conceptual designs. In DAWAK, pages 80–95, 2011.

[18] O. Romero and A. Abell'ó. Automating multidimensional design from ontologies. In ACM DOLAP, pages 1–8, 2007.

[19] E. P. Shironoshita, M. T. Ryan, and M. R. Kabuka, "Cardinality estimation for the optimization of queries on ontologies," SIGMOD Record, vol. 36, pp. 13–18, 2007.

[20] A. Simitsis, D. Skoutas, and M. Castellanos. Natural language reporting for ETL processes. In ACM DOLAP, pages 65–72, 2008.

[21] D. Skoutas and A. Simitsis. Designing ETL processes using semantic web technologies. In ACM DOLAP, pages 67–74, 2006.

[22] M. Stocker, A. Seaborne, A. Bernstein, C. Kiefer, and D. Reynolds, "Sparql basic graph pattern optimization using selectivity estimation," in Proceedings of the 17th international conference on World Wide Web, WWW '08, (New York, NY, USA), pp. 595–604, ACM, 2007. University Science, 1989.

[23] Z. Djilani, N. Berkani, L. Bellatreche : Towards Functional Requirements Analytics, ISOLA, 358-373, 2016.

[24] Z. Djilani and S. Khouri. Understanding user requirements iceberg: Semantic based approach. In MEDI, pages 297–310, 2015.

Construction automatique des espaces de réponse dans les environnements multi-sources

Kamal BAL

Directeurs de thèse : Pr. Youssef Amghar et Pr. A. R. Ghomari

Équipe : MSI

Résumé— Dans des environnements riches en information, les utilisateurs ont besoin d'extraire les informations les plus appropriées parmi une grande masse d'information. La complexité qu'entraîne cette recherche nécessite de fournir aux utilisateurs des outils pour obtenir les informations appropriées à partir de différentes sources d'information et organiser cette information d'une manière compréhensible pour un utilisateur lambda. La prise en compte des relations sémantiques entre les contenus extraits permettrait de construire des espaces de réponses cohérents, diversifiés et bien organisés.

I. CONTEXTE ET PROBLEMATIQUE

La recherche d'information agrégée (*aggregated search*) est un paradigme de recherche d'information qui a été définie pour la première fois dans un atelier à ACM SIGIR¹ 2008 comme étant une tâche cherchant à rassembler des informations provenant de sources différentes, et à les présenter dans une seule interface [1],[2].

Ce paradigme a reçu un grand intérêt dans la recherche et de l'industrie [3]. Un grand nombre d'approches et de systèmes ont été développés [5],[6],[7],[8],[9],[10],[11]. La plus part de ces approches permettent de rechercher des informations dans plusieurs sources et de les fusionner dans une même liste de résultats. Cependant, même s'il est admis que cette façon de présenter les résultats est très intéressante, il reste que ce n'est pas efficace. Ces approches laissent encore à la charge de l'utilisateur la tâche de sélectionner les bouts d'information parmi la liste des résultats et de les organiser pour construire une réponse organisée et cohérente.

L'objectif de notre travail est de proposer une approche de construction automatique des espaces de réponses pertinents, cohérents et bien organisés.

II. ETAT DE L'ART

A. Paradigmes de recherche multi-sources

Un état de l'art relatif à la recherche d'information multi-sources a été réalisé. Les points suivants ont été abordés :

- Introduction et une présentation de l'axe de recherche.
- Différents paradigmes de la RI multi-sources.
- Focalisation sur la RI agrégée.

B. Résultats et discussion sur état de l'art

Etude comparatives des différentes solutions développées dans le cadre de la recherche agrégée présentée dans le cadre de ICEIS'2013 (*15th international conference on enterprise information system*) [4]. De cette étude deux insuffisances majeures ont été identifiées dans ces approches:

- Non prise en compte des relations sémantiques entre les contenus restitués (simple groupement et fusion d'information hétérogènes)
- incohérence des espaces de résultats. aucune approche ne garanti la cohérence des résultats générés.

III. ELEMENTS D'UNE PROPOSITION

Nous avons essayé de proposer une nouvelle approche en nous inspirant du concept de la recherche composite [13,14]. La recherche composite est appliquée souvent dans des environnements semi-structurés où le focus est porté sur seulement la validité des agrégats et de leurs diversités.

A. Formulation du problème :

$S = \{S_1, S_2, S_3, \dots, S_n\}$: l'ensemble des sources

Q : la requête utilisateur.

$R = \{R_1, R_2, \dots, R_m\}$: avec R_i = ensemble des résultats de recherche pour la requête Q dans la source S_i ,

KB : une base de connaissance externe (Wikipedia, Dbpedia ou une ontologie du domaine, ...).

C_{INTER} et C_{INTRA} = deux ensembles de contraintes

Objectif: Construire un espace de réponses RS composé des meilleurs agrégats Agr_i

$RS = \{Agr_1, Agr_2, \dots, Agr_k\}$ tel que :

- Agr_i est composé de plusieurs items résultats.
- Agr_i doit respecter l'ensemble des contraintes intra agrégat C_{INTRA}
- L'espace des réponses RS doit respecter l'ensemble des contraintes inter agrégats C_{INTER} .

Les contraintes Inter-agrégat C_{inter} :

- **La pertinence:** l'agrégat doit être composé d'items résultats pertinents à la requête utilisateur.
- **La cohérence:** l'agrégat doit être cohérent « conceptuellement », ie. composé d'items concernant un seul aspect, facette, sous concept de la requête.
- **La diversité:** l'agrégat doit être composé d'items résultats issus de sources différentes.

Les contraintes Intra-agrégats C_{intra} :

- **La diversité:** Les agrégats formant l'espace de réponse doivent être diversifiés (traitent des aspects, concepts, facette différentes de la requête)..

B. Difficultés

Si la recherche composite dont nous nous inspirons utilise des sources homogènes et souvent structurés, dans notre cas les sources sont hétérogènes. Les défis sont engendrés par les points suivants :

- **Hétérogénéité des items résultats :** Il s'agit de comparer des items résultats hétérogènes (texte, image, vidéo, ...)

¹ ACM Special Interest Group on Information Retrieval

qui ont des représentations différentes. Comment allons-nous les représenter? comment les comparer afin de pouvoir mesurer leurs similarités?

- **Complexité du problème** : il est connu que générer un ensemble optimal d'agrégats sous contraintes est un problème difficile (NP Complet) [13]. Sachant qu'on doit respecter un grand nombre de contraintes (pertinence, diversité, cohérence)

C. Solutions apportées:

Pour palier aux défis cités ci-dessus, nous nous sommes basés sur deux éléments essentiels :

1. **Choix d'une source centrale (noyau)** : Pour diminuer la complexité du problème, nous considérons d'abord une source centrale S_0 . Cette source est supposée contenir du contenu principal, de préférence textuel et bien élaboré. La requête est en premier lieu envoyée vers cette source (Fig. 1). Les résultats de cette interrogation sont ensuite groupés par un Clustering. Une sous requête est générée à partir de chaque Cluster. La sous requête est composée de la requête initiale à laquelle sont ajoutées les concepts les plus fréquents du Cluster. Une recherche exploratoire est ensuite effectuée dans les autres sources pour former les résultats agrégés (Fig. 2).

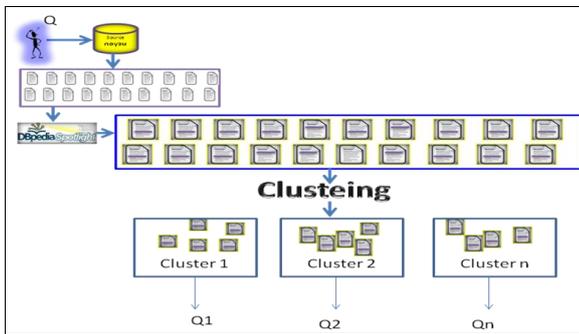


Fig. 1. Interrogation de la source noyau

2. **Utilisation d'une source de connaissance externe** : Une base de connaissance externe est utilisée pour assurer la cohérence des résultats de recherche qui seront regroupés sur la base de leurs relations sémantiques. Cela permettra aussi de palier à la difficulté relative au problème d'hétérogénéité des contenus. Tous les contenus des sources seront représentés à base d'entités (concepts) de la source de connaissance. Cela facilitera la mesure de leurs similarités. Ne seront agrégés ensemble dans un même agrégat que les contenus liés et représentés par les mêmes entités.

IV. DIFFICULTES RENCONTREES

Les difficultés rencontrées sont tout d'abord des difficultés techniques liées à l'implémentation d'un prototype pour l'expérimentation de l'approche proposée. Cette implémentation nécessite la maîtrise d'outils et langages en pour l'implémentation des algorithmes de Clustering et l'interrogation des bases de connaissance. D'autres difficultés liées à la procédure d'évaluation de la notion de cohérence. Si en recherche d'information des mesures d'évaluation de la pertinence existent, la cohérence de l'espace des résultats est un concept difficile à évaluer. Nous prévoyons des expérimentations empiriques pour évaluer cette notion.

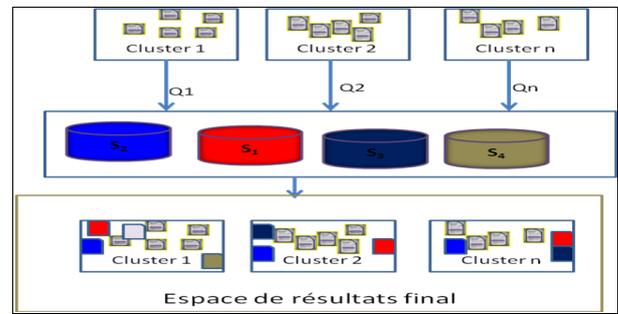


Fig. 2. Exploration des autres sources et génération des agrégats

Références

- [1] V. Murdock and M. Lalmas, editors. SIGIR 2008 Workshop on Aggregated Search, New York, NY, USA, ACM, 1, 3-7. 2008.
- [2] S. Sushmita, M. Lalmas, A. Tombros. Using digest pages to increase user result space: Preliminary designs. Proceedings of the 2008 ACM SIGIR Workshop on Aggregated Search. Singapore, July 2008.
- [3] A. Kopliku, K. M. Boughanem. Aggregated search: potential, issues and evaluation. Technical report: IRIT/RT-2009-4-FR, IRIT, September 2009.
- [4] K. Bal, A.R Ghomari, Y. Amghar, "Aggregated search techniques usability", ICEIS'2013 (15th International Conference on Enterprise Information Systems), ESEO Angers Loire Valley, France, 4-7/07/2013.
- [5] C. Paris, N. Colineau, A. Lampert and K. Vander Linden. Discourse Planning for Information Composition and Delivery: A Reusable Platform. Journal of Natural Language Engineering, 16(1):61-98, Cambridge University Press, 2010.
- [6] H. Elmeleegy, J. Madhavan, A. Y. Halevy, Harvesting Relational Tables from Lists on the Web. Proceedings of the VLDB Endowment (PVLDB), 2009, pp. 1078-1089. 2009.
- [7] C. Paris, S. Wan, P. Thomas : Focused and aggregated search: a perspective from natural language generation. Information Retrieval, Vol. 13, issue 5, 434-459. October 2010.
- [8] C. Sauper, R. Barzilay, Automatically Generating Wikipedia Articles: A Structure-Aware Approach, In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009.
- [9] I. Krichen, A. Kopliku, K. Pinel-Sauvagnat, and M. Boughanem, Une approche de recherche d'attributs pertinents pour l'agrégation d'information. In Proceedings of INFORSID, 385-400, 2011.
- [10] M. Boughanem, J. Savoy., Recherche d'information états des lieux et perspectives, Hermès Science, Avril 2008.
- [11] D. R. Radev, W. Fan, and Z. Zhang. Webinence: A personalized web-based multi-document summarization and recommendation system. In NAACL Workshop on Automatic Summarization, Pittsburgh, PA, 2001.
- [12] S. Sushmita. Study of results presentation and interaction for aggregated search. Phd Thesis, University of Glasgow. 2012.
- [13] S. Amer-Yahia, F. Bonchi, C. Castillo, E. Feuerstein, I. Méndez-Díaz, and P. Zabala. Complexity and algorithms for composite retrieval. In Proceedings of the 22nd international conference on World Wide Web companion, pages 79-80. International World Wide Web Conferences Steering Committee, 2013.
- [14] S. Amer-Yahia, F. Bonchi, C. Castillo, E. Feuerstein, I. Méndez-Díaz, and P. Zabala. Composite retrieval of diverse and complementary bundles. 2013.

La Couverture dans les réseaux de capteurs par les théories de l'incertain

Adda Boualem

Directeurs de thèse : Youcef Dahmani

Équipe de Recherche: /

Résumé—Le problème de la couverture dans les réseaux de capteurs sans fil (WSN) est un problème non trivial. Pour ce fait, il a été largement étudié au cours des dernières années en particulier, lorsqu'est combiné avec d'autres critères comme, la position (distance euclidienne), la connectivité et l'efficacité énergétique. Nous nous intéressons dans notre recherche à tenter d'optimiser ; si possible résoudre deux défis concernant ; (a) le problème de la couverture, à savoir, maximiser la durée de vie du réseau en minimisant la consommation énergétique et (b) garantir la connectivité maximale du réseau de capteurs. Par conséquent, garantir une couverture maximale et fournir une bonne qualité de couverture (QoC) dans le réseau (WSN).

I. Introduction

Les réseaux de capteurs sont dédiés pour une surveillance autonome des phénomènes physique, des catastrophes, le contrôle continu de l'environnement, le contrôle continu des frontières, ainsi les applications médicales et les applications militaire sur une large ou petite échelle.

La couverture est de comment surveiller une zone d'intérêt (RoI) par des ensembles minimaux des nœuds de capteurs ? Cette étape nécessite un équitable déploiement initialement avec une dispersion équitable. Le déploiement influence directement sur la connectivité, la qualité de service et la durée de vie du réseau de capteurs ; de telle sorte qu'un déploiement équitable implique une connectivité garantie, une couverture garantie ; en conséquence, une augmentation de la durée de vie du réseau et une bonne amélioration de la qualité de service (QoS).

La couverture dans les réseaux de capteurs quel que soit de type ; *Barrier Coverage*, *Target Coverage*, ou *Area Coverage* et quel que soit l'application utilisée est un problème NP-Hard à cause de ne pas savoir le nombre minimal des nœuds à déployer sur la zone d'intérêt et dans quelles positions. Un immense recherches à été réalisés pour résoudre ce problème (la couverture dans les réseaux de capteurs sans fil). Parmi ces travaux celles qui basent sur les méthodes mathématique (la géométrie ; comme les cercles, les triangles, les Matrices et les losanges), et celles basent sur la théorie du graphe (Le diagramme de Voronoi et le diagramme de Deloney) et celles

qui utilisent les méthodes de la recherche opérationnel (les techniques de l'ordonnancement)..., mais quelques travaux basant sur les méthodes statistiques, et les théories de l'incertain.

Le problème de la couverture dans les réseaux de capteurs sans fils concentre sur la réponse de la question suivante « comment font les capteurs pour assurer la couverture optimale de l'espace physique ? ». Ce qui veut dire la qualité de surveillance qui peut être fournit par WSN « comment améliorer la qualité de surveillance en utilisant les réseaux de capteurs ? ». Par conséquent, maximiser la couverture on prend en considération les contraintes par les quelles souffre les nœuds de capture ; comme la contrainte de batterie limitée, la capacité de traitement et de mémorisation limitée, le rayon de communication, et le rayon de surveillance très limités.

II. Problématique

Le problème de la couverture dans les réseaux de capteurs sans fil (WSN) est un problème non trivial (NP-Hard). Pour ce fait, il a été largement étudié au cours des dernières années, en particulier, lorsque est combiné avec d'autres critères comme ; la *position* (distance euclidienne), la *connectivité* et *l'efficacité énergétique*. Les méthodes utilisées pour étudier la couverture sont classées comme montrée la figure (Fig. 1). La couverture est classée sur trois catégories ; chaque catégorie peut regroupe des problèmes de couverture de mêmes objectives ; le premier type est : *couverture de cibles (target coverage)*, *couverture par Barrière (Barrier Coverage)* et *Couverture de surface (area coverage)*. Chaque type de couverture selon la nécessité peut utiliser un seul nœud capteur comme (*I-Couverture ou couverture simple*), comme peut utiliser plusieurs comme (*K-couverture*). Nous nous intéressons dans notre recherche à tenter de résoudre deux défis concernant ; (a) le problème de la couverture de zone, à savoir, maximiser la durée de vie du réseau en minimisant

la consommation énergétique et (b) garantir la connectivité maximale du réseau de capteurs. Par conséquent garantir une couverture maximale et fournir une bonne qualité de couverture (QoC).

III. Travaux Réalisés

Les travaux que nous avons réalisés depuis mon dernière inscription sont résumés dans le tableau suivant :

Titre du papier	Journal	Conférence .International	Etat
The Redeployment for Area Coverage Maximization in WSN Based on Descriptive Statistics Parameters Decision		ICNAS 2017 ANNABA et dans ICT-DM2017	Accepté dans les deux
Energetic Sleep- Scheduling via Probabilistic Interference K-Barrier Coverage with Truth-Table Technique in Sensor Network		ICMIT_IEEE Adrar	Accepté
Area Coverage Optimization in Wireless Sensor Network by Semi-Random Deployment		SENSORNETS (6 th international Conference on Sensor Networks)	En traitement
Area Coverage and Efficient-Energy Based on Voronoi Diagram, Clustering and minimal dominating sets in WSNs	Journal of Communications and Networks		En traitement
Dempster-Shafer Theory-based Area Coverage in Wireless Sensors Network	Wireless Networks		En traitement
Semi-Random deployment based Area Coverage Optimization in Wireless Sensor Network	AD HOC & SENSOR WIRELESS NETWORKS		En traitement

IV. Les Travaux à envisager à faire pendant l'année courante

Dans l'année prochaine nous allons rectifier les travaux refusés de l'année précédente. En remédiant les points faibles de ces deniers travaux et rédiger un nouveau papier concernant la couverture de surface dans les réseaux de capteurs sans fils, en suivant la méthodologie suivante :

- Exposer les différentes difficultés, les points faibles et les points forts des approches proposées.
- Proposer des améliorations pour remédier aux faiblesses de ces travaux.
- Proposer de nouvelles idées pour une nouvelle approche et rédiger un autre papier.

Pour but d'obtenir des publications.

References

- [1] Carlos E. Otero et "Analysis of Wireless Sensor Networks Deployments using Vertical Variance Trimming and the Analytic Hierarchy Process" International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.2, No.1, March 2011
- [2] A. BOUALEM, Y. DAHMANI, and A. MAATOUG, "PROBABILISTIC INTERFERENCE K-BARRIER COVERAGE WITH TRUTH-TABLE TECHNIQUE FOR BORDER SECURITY IN WIRELESS SENSOR NETWORKS (PIBC-TTT)", ICWSN 2014 London, United Kingdom: 16th International Conference on Wireless Sensor Networks, OCTOBER 2014.
- [3] A. BOUALEM, Y. DAHMANI, and A. MAATOUG, « The Combination Operator of Information Sources by a New Expressive Matrix », Journal of Geographic Information System, (JGis), 20 AOUT 2015.
- [4] A. BOUALEM , Y. DAHMANI, and A. MAATOUG: « The Redeployment for Area Coverage Maximization in WSN Based on Descriptive Statistics Parameters Decision », 3rd International Conference on Networking and Advanced Systems 13-14 December 2017 | Badji Mokhtar University, Annaba, Algeria.
- [5] A. BOUALEM , Y. DAHMANI, and A. MAATOUG: « Energetic Sleep- Scheduling via Probabilistic Interference K-Barrier Coverage with Truth-Table Technique in Sensor Network », International Conference on Mathematics and Information Technology, Adrar 04-05 December, 2017.

Location privacy in the Internet of things

Amina ZEMMACHE

Directeurs de thèse : Djamel Tandjaoui, Souad Benmezziane

Équipe : DSI

Abstract— Location information is core context data needed to provide tailored Internet of Things services and applications. However, location data may be mined and used to infer more personal information such as user's motion patterns, gender, hobbies, etc. This may result in the users getting spam mail, advertisements, price discrimination or cause other more serious issues that can disturb the user's normal life. Thus, the main objective of our work is to propose an effective algorithm to protect user's location and trajectory privacy by considering both computational costs and various privacy requirements.

I. INTRODUCTION

The Internet of Things (IoT) connects everyday things comprising a vast array of sensors, actuators, and smart devices. These things also known as smart things will be integrated into streets, cars, homes, buildings, and other public environments. They will also be carried by people and communicate with each other locally or with IoT applications remotely over the Internet. Therefore, things will have the capabilities of automatically sensing and communicating user's contextual and private information. Thus, it provides a huge opportunity for hackers to compromise security and privacy of users in general, and location privacy in particular. User must ensure that he can retain his location anonymity in this super connected world.

Over the last two decades, a significant effort on location privacy protection has been expended by the research community [1-3]. Most of them are based on the k -anonymity cloaking technique, which was first proposed in [4]. The k -anonymity cloaking technique uses a trusted third-party to select region containing user's real position and $k-1$ other users. However, the heavy reliance on a third party leads to a single point failure and also once a TTP is compromised, user location information would likely be exposed. To preserve user's location privacy without using any third party, the dummy location techniques was proposed [5].

II. PROBLEM STATEMENT

Although many techniques have been proposed to deal with location privacy problem [1-5], some limitations still remain and they cannot be applied to the Internet of things. We can summarize these limitations as follows: (i) Most of the existing location privacy mechanisms have not been addressed with respect to emerging IoT devices which have limited capabilities in terms of energy, computational power and memory. (ii) Existing privacy preserving techniques anonymize location information so that adversaries cannot relate location information to specific users. However, these techniques do not take into account the case where a user

requirements to adjust the degree of the disclosed information based on the context of their use (e.g., based on the time, the current location of the user or type of service). (iii) Existing location privacy techniques don't consider attacker's background information. Thus, it is hard to effectively guarantee the desired k -anonymity. Thus, the challenge is to propose a lightweight and effective privacy protection scheme which adaptively modifies its parameters, in order to guarantee personalized location-privacy protection requirements against adversaries.

III. WORK ACHEIVED

We are particularly interested in dummy-based anonymization mechanism for location protection [5], which is one of the most typical techniques to preserve location privacy without relying on a trusted third party (TTP). The device sends his true position data along with $k-1$ false location data when requesting for service, so service provider cannot distinguish the user's real location from the $k-1$ locations. We can distinguish two problems in dummy location selection: (i) location privacy protection for single query, [6, 7] i.e. safeguard user's actual location/position. (ii) Protection trajectory privacy for continuous queries [8], which aim to protect the path that user is continuously monitoring for certain period of time.

We argue that existing dummy location selection algorithms don't consider the following criteria:

- **Attacker's background knowledge:** The adversary may have prior knowledge about the trajectories that may be possible in the geographic space. More precisely, the adversary knows the mobility pattern of the users in the considered space. He can model the users' movement, predict future position and subsequently differentiate trajectories of true users from the dummies.
- **Semantic of locations:** most of existing dummy location techniques divides a space on grid cells or circles without considering the semantic of each location. Indeed, when all dummies locations and the real one are of the same type (e.g. hospital, clinic), the adversary can infer what the user is doing with a big probability (e.g. user have some health problems).
- **Adaptive privacy or user-centric privacy:** The existing dummy-based location privacy-enabling approaches are static i.e. they do not consider heterogeneous user privacy requirements.
- **Time complexity and energy consumption:** the cost of existing algorithms increases exponentially with the

growth of the value of k . However, IoT devices are resource constrained (i.e. limited battery power, small memory unit and low processing capability) and thus algorithm must be lightweight enough to be realized on IoT devices.

Our challenge is to conceive a lightweight user-centric algorithm of dummy locations selection which overcomes the aforementioned issues.

In the other hand, we have also performed an exploratory survey about the proposed strategies to preserve location privacy at physical layer. Indeed, the open nature of wireless network environments paves the way to passive attackers to locate and track the thing's movements exploiting well-known radio positioning schemes (e.g. trilateration) by measuring received signal strength (RSS) or others signal parameters. Physical layer location privacy has not gained a lot of attention in literature. Until now, very few schemes were proposed. These schemes are limited to power randomization [9-12] that uses different power levels in transmitting packets, noise addition [13-17] that attempt to decrease the accuracy of sender's localization to noise ratio and smart antenna [18] that attempt to direct radio towards the intended receivers using beamforming techniques.

IV. FUTURE WORKS

For the future work, we will follow these steps:

- Propose solution to enhance and overcome the drawbacks of existing dummy location privacy algorithms in the context of IoT.
- Implement and evaluate performance of proposed solution.
- Submit a survey paper about location privacy techniques at physical layer.

V. CONCLUSION

We have presented in this short paper our achievement during this year. We have identified weaknesses of existing dummy location privacy protection algorithms. The next challenge is to find effective solution to overcome these weaknesses and which adapts to the context of Internet of things

References

[1] C. Chow, M. Mokbel, et al. "Casper*: Query processing for location services without compromising privacy." ACM, 2009.
 [2] P. Kalnis, et al. "Preventing location-based identity inference in anonymous spatial queries." IEEE Transactions on Knowledge and Data Engineering, 2007.
 [3] B. Gedik, L. Liu. "Protecting Location Privacy with Personalized k -

Anonymity: Architecture and Algorithms" IEEE Transactions on Mobile Computing, 2008.
 [4] M. Gruteser and D. Grunwald. "Anonymous usage of location-based services through spatial and temporal cloaking." ACM Mobisys, 2003.
 [5] H. Kido, et al. "An anonymous communication technique using dummies for location-based services." IEEE, 2005.
 [6] Lu, Hua, et al. "Pad: privacy-area aware, dummy-based location privacy in mobile services." Proceedings of the Seventh ACM International Workshop, 2008.
 [7] N. Ben, et al. "Achieving k -anonymity in privacy-aware LBS." INFOCOM, 2014 Proceedings IEEE, 2014.
 [8] Liao et al. "The framework and algorithm for preserving user trajectory while using location-based services in IoT-cloud systems." Cluster Comp, 2017.
 [9] T. Jiang, H. et al. "Preserving location privacy in wireless LANs," in Proc. 5th Int. Conf. MobiSys, 2007.
 [10] R. El-Badry, et al., "Hyberloc: Providing physical layer location privacy in hybrid sensor networks," in Proc. IEEE ICC, Cape Town, South Africa, 2010.
 [11] S. Taha and X. Shen, "A physical-layer location privacy-preserving scheme for mobile public hotspots in NEMO-based VANETs," IEEE Trans. Intell, 2013.
 [12] Yao, Lan et al. "Pattern regulator: Protecting temporal usage privacy for wireless body area sensor networks" Distributed Computing Systems Workshops (ICDCSW), IEEE 33rd International Conference, 2013.
 [13] El-Badry, Rania, et al. "Hidden anchor: Providing physical layer location privacy in hybrid wireless sensor networks." New Technologies, Mobility and Security (NTMS), 3rd International Conference on. IEEE, 2009.
 [14] R. El-Badry et al., "Hidden anchor: A lightweight approach for physical layer location privacy," J. Comput. Commun, 2010.
 [15] Oh, Sangho, et al. "Multi-node coordinated jamming for location privacy protection." military communications conference, 2011.
 [16] Oh, Sangho, et al. "Phantom: Physical layer cooperation for location privacy protection," Proceedings IEEE, 2012.
 [17] Lan Yao et al. "Virtual Node Based Location Privacy Protection in Wireless Body Sensor Networks", IEEE, 2015.
 [18] T. Wang et al, "Location privacy protection from rssi localization system using antenna pattern synthesis," in Proc. IEEE INFOCOM, Shanghai, China, 2011.

Design of a system for the diagnosis of diseases based on the clustering of genomic data

PhD student : Billel Kenidra

Thesis director : Mohamed Benmohammed

Abstract— While many algorithms have been evolved and applied the clustering process of high-dimensional datasets, the running time still remains an enormous issue. In practice, the conventional clustering techniques require large amounts of computational time when it comes to process high-throughput genomic datasets with high-dimension attributes. To cope with this problem, we have established a very fast and accurate method for big genomic data. The proposed method has been shown to perform very well in experimental datasets regarding running time as well as accuracy.

I. INTRODUCTION

Bioinformatics is the science of managing, analyzing, extracting, and interpreting information from biological sequences and molecules. Recent advancements in microarray technology allow simultaneous monitoring of the expression levels of a large number of genes over different experiment conditions. Facing this huge amount of data, the biologist cannot simply use the traditional techniques in biology to analyze the data.

In fact, information technologies are needed. Cluster analysis is of considerable interest and importance in the field of bioinformatics, either by clustering the genes or by clustering experiment conditions (samples). The clustering of genes is used to identify groups of genes with similar patterns of expression, aiming at helping to answer questions of how gene expression is affected by various diseases and which genes are responsible for specific diseases. The clustering of samples is used to organize the samples into intrinsic clusters such that samples with high similarity belong to same cluster. The significance of this clustering assists in diagnosis of the disease condition, and it discloses the effect of certain treatment on genes. [1].

II. PROBLEMATIC

Since biologists inherently need effective and efficient computational methods to interpret the vast amounts of data that are constantly collected in genomic research, into knowledge, bioinformatics has emerged to meet this need. Bioinformatics is likely to be as a key to decipher encoded information in vast genomic structures.

Going through high-throughput datasets with high-dimension attributes in order to identify interesting biological features at a reasonable time, is a great challenge. DNA-Microarray technology simultaneously monitors the expression profiles of thousands of genes over various experimental conditions. Identifying co-expressed genes and coherent patterns is the central goal of clustering process, and it is an important task in bioinformatics as it helps biologists to

gain insights on gene functions, because genes with similar functions exhibit similar expression patterns [3].

Data-clustering mainly aims in forming the amount of unmanaged data to manageable piles, by discovering homogeneous groups. It is a combinatorial problem, because the number of partitions that can be obtained grows exponentially with the volume of data to be classified and the number of clusters [2].

The clustering process is used to identify cancer subtypes based on gene expression and DNA methylation datasets, as information on cancer subtypes is critically important for understanding tumor heterogeneity, detecting previously unknown clusters of biological samples, which are usually associated with unknown types of cancer will, in turn, provides the opportunity to prescribe more effective treatments for patients, as various subtypes of cancer often respond disparately to the same treatment. While DNA methylation database is a kind of extremely large-scale datasets, running time still remains a major challenge. In fact, traditional clustering algorithms are too slow to handle high-dimensional biological datasets, they usually require large amounts of computational time [3].

III. ACHIEVED WORKS

« A data-clustering approach based on artificial ant colonies with control of emergence combined with K-Means approach » ACM conference, published in 2015.

« A partitional approach for genomic-data clustering combined with K-Means algorithm » IEEE conference, published in 2016.

« An Ultra-Fast Method for Clustering of Big Genomic Data » IGI Global International Journal of Applied Metaheuristic Computing, under reviewing since July 2017.

Due to the inability of conventional methods to deal with big genomic datasets in efficient manner within tolerable time limits, a novel method is proposed in this last paper. This method could be used as a way to ultra-quickly cluster high-throughput genomic datasets with high-dimension attributes, this innovative method has been found capable of addressing the running-time challenge.

Another plus of our method is the accuracy, in spite of the fact that high dimensionality increases the difficulty and the computational cost to attain the optimal solution, the proposed method not only shortens the amount of time needed for clustering, it is also able to process any amount of data with high accuracy, whereas the previous approaches do not provide that.

IV. PERSPECTIVES

I plan to draw the chapters and start writing the dissertation, while waiting the reviewers replies of the submitted journal...

V. CONCLUSION

The clustering of genomic datasets mainly aims at exploring the genetic relationships of deadly diseases and understanding the cell behavior. Cancer clustering analysis based on genomic data is a highly important topic for the cancer study, because it reveals biological insight into unknown functions of genes, since genes that are in the same cluster, are likely to be sharing the same function.

While many algorithms have been evolved and applied the clustering process of high-dimensional datasets, the running time still remains an enormous issue. In practice, the conventional clustering techniques require large amounts of computational time when it comes to process high-throughput genomic datasets with high-dimension attributes.

To cope with this problem, we have established a very fast and accurate method for big genomic data. The proposed method has been shown to perform very well in experimental datasets regarding running time as well as accuracy. Finally, we are in high hopes that the proposed technique would be of immense help in facilitating early diagnosis and effective prognosis of cancer.

References

- [1] B. Kenidra, and M. Benmohammed, "A Partitional Approach for Genomic-Data Clustering Combined with K-Means Algorithm," CSE/EUC/DCABES 2016 Paris, IEEE conference, pp. 114-121, August 2016.
- [2] B. Kenidra, and M. Benmohammed, "A data-clustering approach based on artificial ant colonies with control of emergence combined with K-Means approach," IPAC 2015 Batna, ACM conference, doi:10.1145/2816839.2816931, November 2015.
- [3] B. Kenidra, and M. Benmohammed, "An Ultra-Fast Method for Clustering of Big Genomic Data," IGI Global International Journal of Applied Metaheuristic Computing, under reviewing since July 2017.

Une approche multidimensionnelle pour améliorer les systèmes de recommandation

Candidat : Kouadria Abderrahmane

Encadreur : Dr. Omar Nouali

Abstract— Recommender systems are effective tools for guiding Web users in a very big medium of data. However, the main weakness of these systems lies in the sparsity of the user-item matrix which in turn degrades the performance. For solving this, we propose a hybrid recommender system, which uses the users' information and the items' information. This system is hybridization between two different systems using two different information for each one of them. The first level hybridizes user-based and item-based collaborative recommender systems and user demographic recommender system and item demographic recommender system. The second level hybridizes the two resulting systems. Many weighting approaches are tired out. The results show the effectiveness of this system as it reduces the error.

I. INTRODUCTION

Les systèmes de recommandations représentent un thème de recherche fondamental en plein essor, à l'intersection entre plusieurs grandes disciplines telles que l'apprentissage automatique, les systèmes multi-agents et les sciences cognitives. Ils constituent en outre un cadre applicatif ambitieux pour la communauté des chercheurs en Intelligence Artificielle de par leur très grande complexité et les nombreuses contraintes qu'ils génèrent.

Le but de ces systèmes est d'améliorer les interactions entre le grand public et les systèmes de recherche et d'accès à l'information. Il est en effet devenu difficile, dans le contexte de masse de données hétérogènes que constitue Internet, d'identifier les items les plus pertinents. Il s'agit d'assister les utilisateurs dans leurs recherches, mais également de leurs proposer des items susceptibles de les intéresser mais qu'ils n'auraient pas consulté spontanément.

II. PROBLEMATIQUE

Malgré leur popularité croissante les systèmes de recommandation connaissent plusieurs problèmes :

Masse critique : Cet aspect illustre la difficulté à gérer du fait qu'il existe peu d'articles effectivement évalués, ou peu d'utilisateurs qui procèdent à ces évaluations. A cet effet, le système de recommandation ne possède que peu ou pas de données pour générer ces prédictions.

Démarrage à froid : Souvent, on se retrouve confronté au problème qu'un utilisateur ne soit comparable avec aucun autre. Ce problème est dû au fait que peu ou pas d'utilisateurs ont évalué un article donné, ou qu'un utilisateur donné a évalué très peu ou pas d'articles.

Principe d'induction : Les systèmes de recommandation se basent sur le principe qu'un utilisateur qui a exhibé un comportement dans le passé tendra à exhiber un comportement semblable dans le futur. Cependant, ce principe n'est pas nécessairement valable dans le contexte réel. En effet,

un utilisateur peut changer complètement de domaine d'intérêt ou en avoir plusieurs.

Architecture du système : Les systèmes de recommandation, tels *qu'Amazon*, sont typiquement basés sur une architecture centralisée. Où le moteur de recommandation centralisé permet de sauvegarder les profils d'utilisateurs et le calcul des recommandations dans un serveur central. Cependant, en dépit de leur popularité, les systèmes de recommandation centralisés souffrent de plusieurs problèmes: le coût, la robustesse, la sécurité, la mobilité, etc.

Mobilité : Pour pouvoir avoir accès à des recommandations à tout moment et en tout endroit (épicerie, film, musique, électronique, etc.) en utilisant son cellulaire, PDA, PALM, etc.

Sécurité et crédibilité : Les systèmes de recommandation ne peuvent pas empêcher les actes de tromperie. Il devient ainsi facile de se forger une nouvelle identité et de se livrer au vandalisme, comme fournir de mauvaises informations au système.

Explication de la recommandation : Il faut présenter la recommandation pour qu'elle soit acceptée par l'utilisateur et pour améliorer les recommandations futures. Dans ce but, il est possible d'utiliser, comme, une explication sous forme de texte ou de graphiques.

Collecte des préférences : Une des étapes les plus importantes et les plus difficiles des systèmes de recommandation est la collecte des préférences des utilisateurs. En effet, l'obtention des évaluations de la part des utilisateurs sur une ressource donnée qui leur a plu, moins plu, ou pas du tout plus, est une tâche ardue.

Complexité computationnelle : L'évaluation des similarités, une partie intégrale du filtrage de collaboration, implique quelques processus comportant de nombreux calculs. Pour un nombre important d'utilisateurs, le calcul de la similarité pour tous les individus du voisinage devient infaisable. Ainsi, des mécanismes intelligents de filtrage sont nécessaires ne sacrifiant pas trop d'information importante.

Protection de la vie privée : Un autre problème qui touche les systèmes de recommandation est la protection des informations sensibles constituant le profil utilisateur (information personnelle, intérêts, goûts, habitudes, etc.). Vu la nature de l'information, ces systèmes doivent assurer une telle protection. Ainsi, des moyens, pour préserver l'anonymat des utilisateurs et chiffrer les données transmises, sont nécessaires.

L'objectif principal de cette recherche est d'améliorer les systèmes de recommandation en proposant une approche qui permet d'intégrer de nouvelles dimensions (Sociale, contexte, annotation, traces, relation, profils,...etc.) dans le processus de recommandation. Il s'agit à la fois d'élaborer un modèle de recommandation permettant une prédiction précise, robuste et rapide, et ainsi de valider le modèle sur des données réelles, en termes de performance de prédiction.

III. RESUME DES TRAVAUX REALISES :

A. Etat de l'art sur :

- 1- Systèmes de filtrage d'information et systèmes de recommandation ;
- 2- Systèmes de recommandation multicritère et multidimensionnel ;
- 3- Systèmes de recommandations sensibles au contexte (CARS). Ici nous avons détaillé les différents types de données contextuelles dont on peut disposer et les différentes approches de contextualisation existantes. Aussi nous sommes intéressés à la contextualisation sociale d'un problème de recommandation à grande échelle dans le domaine des réseaux sociaux en ligne, pour lequel nous avons présentés les techniques de recommandation basée sur un réseau social.

B. Les solutions proposées :

1. Amélioration du modèle de filtrage collaboratif hybride relatif au système de recommandation

Les systèmes de recommandation sont des techniques utiles pour résoudre le problème de la surcharge d'information. Le Filtrage Collaboratif (FC) représente l'une des techniques de recommandation les plus populaires, dont la plupart de leurs méthodes fondent leur approche sur la prédiction de notes pour générer les recommandations. Beaucoup d'algorithmes récents ont été développés pour améliorer les divers aspects du filtrage collaboratif, cependant, la plupart d'entre eux ne prennent pas en compte les données communes des utilisateurs et les informations des items ou leurs caractéristiques. Dans cet article, nous proposons une nouvelle amélioration du système de recommandation basé sur le modèle de filtrage collaboratif hybride qui combine entre l'algorithme de FC basé-item et l'algorithme de FC basé-utilisateur, dont les similarités entre les items et les utilisateurs ont été modifiés et les évaluations communes prennent totalement l'avantage de la matrice utilisateur-item, pour modifier le calcul de la similarité et la prédiction des notes. Les résultats expérimentaux montrent la supériorité de notre proposition par rapport aux autres approches de l'état de l'art [1].

2. A multi-criteria collaborative filtering recommender system using learning-to-rank and rank aggregation

Finding a top-N list for a multi-criteria recommendation is a challenging problem as we have many criterions for each item. The simplest way seems to find the average; however, this requires finding a score from each criterion and hence compensation effect will occur. Moreover, this seems as we have many recommendation systems working in parallel. This paper takes a new direction by proposing a three-step hybrid system for finding the top-N list for the multi-criteria recommendation system. The first step decomposes the multi-criteria user-item matrix while the second step finds partial-ranked lists for each item using ListRank-FM as a learn-to-rank method. The final step aggregates the partial-ranked lists into a global ranking list using Borda count as a ranking aggregation method. This will reduce the processing time and improve the recommendation quality by representing better the user preference for each criterion. The experimental results on Yahoo!Movie dataset show that the proposed method outperforms both the item-based collaborative recommendation and ListRank-MF [2].

3. Two-level hybridization between demographic and collaborative recommender systems

Hybrid recommender systems usually come up with the best of each individual system. In this direction, many attempts for hybrid systems have been discussed and proposed in the literature. However, no one tries to exploit all facets of the information taken explicitly or implicitly from the user. The work of this paper tries to exploit the demographic information for both users and items besides the set of ratings for user-based and item-based collaborative recommender systems. Hence, four parallel recommender systems will be there in the same time to form two-level hybridization. The first level of the hybridization will be between each two homogeneous systems, i.e. the demographic recommender systems of the users and the items and similarly the collaborative recommender systems of the users and items. The second level of the hybridization will be heterogeneous between the first level hybrid demographic recommender system and the first level hybrid collaborative recommender system. The results show that both hybrid systems of level one and level two are far better than individual ones [3].

IV. CONCLUSION ET PERSPECTIVES :

Le filtrage collaboratif (FC), qui constitue une des solutions techniques implémentant l'intelligence collective, est déjà utilisé avec succès dans plusieurs systèmes de recommandation. Actuellement, le FC est considéré comme la technique de recommandation la plus réussite. Cependant, le filtrage collaboratif présente certaines limitations, comme le cas du problème de la rareté, le problème d'évolutivité et le problème de démarrage à froid. Nos contributions au niveau de cette thèse présentent des solutions pour ces défaillances très souvent rencontrées dans le filtrage collaboratif.

Généralement, dans les systèmes de tagging social, les utilisateurs ne fournissent pas de notation explicite sur les ressources (items) qui les intéressent. Au lieu d'évaluer les items par des notes, les utilisateurs annotent les ressources à l'aide de mots clés appelés tags, qui indiquent l'intérêt et la préférence de l'utilisateur. L'approche qu'on doit utiliser analysera le comportement de marquage (tagging) de l'utilisateur et tentera d'estimer l'intérêt et la préférence de l'utilisateur. Le but de ce travail sera l'utilisation d'un système de recommandation basé sur les données de marquage c'est-à-dire l'exploitation des informations d'annotations (tags) collaboratives fournies par les utilisateurs.

BIBLIOGRAPHIE:

- [1] K. Abderrahmane, N. omar (2014). *Amélioration du modèle de filtrage collaboratif hybride relatif au système de recommandation*. Les 4èmes Journées Doctorales en Informatique de Guelma (JDI'14).
- [2] K. Abderrahmane, N. omar et M. Y. H. Al-Shamri. *A multi-criteria collaborative filtering recommender system using learning-to-rank and rank aggregation*. Information retrieval journal springer (Attente réponse)
- [3] K. Abderrahmane, M. Y. H. Al-Shamri et N. omar, *Two-level hybridization between demographic and collaborative recommender systems*, envoyer à First international conference on Embedded & Distributed Systems, EDiS'2017 (Attente réponse).

Approches réalistes pour le déploiement et le contrôle des réseaux de capteurs sans fil

Doctorant : Mostefa ZAFER

Directeurs de thèse : Mohamed AISSANI

Co-directeur de thèse : Mustapha Réda SENOUCI

Équipe de Recherche : LIA/EMP

Résumé—Le présent travail est un point de situation sur les travaux jusqu'à maintenant réalisés dans le cadre de la préparation de la thèse de doctorat. Il comporte : (1) Un état de l'art sur les différentes solutions proposées dans le domaine considéré réalisé au cours de la 1^{ère} année ; (2) Une description de l'approche proposée, conçue, implémentée et évaluée pendant la 2^{ème} année ; (3) Les travaux planifiés pour la 3^{ème} année.

I. INTRODUCTION

Le réseau de capteurs sans fil (RCSF) est un type particulier des réseaux ad hoc, composé d'un ensemble de capteurs déployés sur une zone géographique pour surveiller ou contrôler un évènement cible. Chaque capteur est alimenté par une batterie de durée de vie limitée et dispose d'une capacité de détection et de traitement ainsi qu'une interface de communication sans fil ; il est capable ainsi de prendre des mesures relatives à l'évènement cible et les communiquer à un point de collecte, dénommé "puits" et ce, selon un routage multi-saut. Pour que ce fonctionnement soit robuste et fiable, les conditions de couverture et de connectivité devant être satisfaites pendant toute la durée de fonctionnement du réseau. La couverture signifie, dans sa définition la plus simple, que chaque point de la zone d'intérêt est couvert par au moins un capteur, cela donne au réseau la possibilité de détecter la survenue de l'évènement cible à n'importe quel point. La connectivité signifie la présence, entre n'importe quel capteur et le puits, d'au moins un chemin, via lequel les données captées sont acheminées.

Pour respecter ces deux conditions (couverture et connectivité), il est nécessaire que les capteurs soient bien répartis sur la totalité de la zone d'intérêt. La solution fréquemment utilisée consiste à déployer les capteurs de manière aléatoire, mais assez dense. Ce déploiement, dénommé aléatoire, est fortement recommandé pour certains cas d'utilisation, notamment lorsqu'il s'agit d'une zone d'intérêt inaccessible, néanmoins, il rend la construction et la maintenance de la topologie logique plus difficiles et ce, à cause du nombre important des capteurs déployés. Aussi, pour les terrains tridimensionnels, ayant généralement des topographies complexes, ce déploiement produit des vides de couverture, qui sont des parties de la zone d'intérêt ayant une très faible densité des capteurs, et par conséquent, un niveau de couverture intolérable, qui altère le fonctionnement globale du réseau. Les limites du déploiement aléatoire sur les terrains tridimensionnels, notamment en termes de qualité de couverture, ont été mises en évidence dans plusieurs travaux de recherche [1] [10].

En effet, la couverture d'un terrain tridimensionnel par RCSF présente plusieurs défis, qui exigent de procéder au déploiement des capteurs d'une manière déterministe, à des

positions calculées au préalable, en fonction de plusieurs facteurs, comportant entre autres la nature de l'application, les types des capteurs à déployer, et les caractéristiques du terrain notamment sa topographie. Ce dernier influe profondément les opérations de capture et de communication ; il est considéré par conséquent un facteur prépondérant à prendre en compte pour obtenir des meilleures performances en termes de couverture et de connectivité.

Ce problème, démontré NP-Complet [1] [10], a fait l'objet de plusieurs travaux de recherche, qui ont proposé, sous différentes formulations, des solutions approximatives basées sur des heuristiques ou des méta-heuristiques. Dans les formulations proposées, le terrain a été représenté par des modèles mathématiques ou numériques et son impact sur les performances du réseau a été pris en compte en introduisant le facteur de visibilité, dont son influence sur les opérations de détection et de communication a été modélisée d'une façon binaire ou probabiliste. L'inconvénient commun de ces solutions réside dans le fait qu'elles traitent la totalité du terrain de la même manière, sans tenir compte des différences de la complexité topographique entre les différentes régions du terrain. A titre d'exemple, la complexité d'une région du terrain est l'un des facteurs primordiaux, à prendre en considération pour estimer le nombre des capteurs à déployer dans cette région, chose qui n'est pas été pris en compte d'une manière claire dans ces travaux antérieurs.

Dans le cadre de ce travail, nous proposons une nouvelle approche de déploiement d'un RCSF sur un terrain 3D, représenté par un modèle numérique triangulaire, dans le but d'assurer la couverture et la connectivité. Cette approche comporte trois phases. **La 1^{ère} phase** consiste à partitionner le terrain en régions de topographies relativement simples, en appliquant une heuristique conçue à cet effet, basée sur l'analyse de visibilité entre les triangles constituant le modèle numérique. Dans chaque région construite, le facteur de visibilité est moins posé et son influence sur la détection et la communication est fortement dégradée. **La 2^{ème} phase** est destinée au déploiement des capteurs sur chacune de ces régions, sans tenir compte du facteur de visibilité, en utilisant des simples heuristiques, dans le but d'assurer la couverture de chacune de ces régions. **La 3^{ème} phase** est destinée à consolider la connectivité globale du réseau, par le recours à l'utilisation des relais devant assurer la communication entre les régions. Ces derniers qui n'ont aucune tâche de détection, sont utilisés dans des solutions antérieures pour assurer la fluidité des données en construisant des clusters. Dans ce travail, nous donnons juste une indication sur notre vision pour leur placement, dont les détails feront l'objet d'un travail ultérieur. En effet, nous envisageons de placer les relais à des positions ayant une large visibilité, situant forcément sur

les frontières entre les zones construites, pour qu'ils jouent le rôle de passerelles interzones. Leur nombre et leurs positions dépendent de plusieurs paramètres à savoir : le nombre de régions construites, le nombre et les positions des puits, ainsi que l'architecture du RCSF envisagée, qui peut être 1-tiers ou 2-tiers.

II. DEPLOIEMENT D'UN RCSF SUR UN TERRAIN 3D ETAT DE L'ART

Dans cette section, nous recensons les différentes approches de déploiement déterministe d'un RCSF sur un terrain tridimensionnel, en mettant en évidence la différence entre elles en termes d'hypothèses, de formulations et de méthodes de résolution utilisées.

La première différence entre ces approches réside dans la manière de modéliser le terrain. Certains travaux utilisent le modèle mathématique, qui représente la surface du terrain par une formule mathématique et donne ainsi les coordonnées de chaque point sur le terrain, sans prendre des mesures réelles. Dans la plupart des approches, le terrain est représenté sous forme d'une matrice, dont la valeur de chaque cellule correspond à l'altitude de l'emplacement qu'elle représente sur le terrain. Ce modèle, dénommé numérique matriciel ou "raster", peut aussi être enrichi par des méthodes de triangulation pour reconstruire et représenter la surface du terrain la sous forme d'un réseau de triangles [41]. Ce dernier modèle est dénommé modèle numérique triangulaire.

Les approches de déploiement déterministe se différencient également dans le modèle de couverture adopté. Ce dernier, qui peut être binaire, probabiliste ou hybride, mesure la qualité de couverture d'un point du terrain par un capteur, en prenant en considération une liste de paramètres, qui se varie d'une solution à l'autre, et qui comporte entre autres la distance capteur-point et l'orientation du capteur, dans le cas où il est supposé directionnel. Le paramètre de la topographie est aussi pris en compte via la condition de visibilité capteur-point qui est vérifiée en utilisant des algorithmes conçus à cet effet [46] [49].

Ces approches visent également des objectifs différents comme la couverture, la connectivité, la discrétion des capteurs, la durée de vie du réseau et le coût de déploiement. Une autre différence entre ces approches réside dans les méthodes utilisées (heuristiques et méta-heuristiques) pour résoudre ce problème une fois formulé en fonction des facteurs et des objectifs cités-supra.

1. APPROCHES DE DEPLOIEMENT BASEES SUR UN MODELE MATHEMATIQUE DU TERRAIN

Dans [11] [20], le terrain est modélisé mathématiquement sous forme d'un cône et les capteurs sont supposés homogènes, omnidirectionnels et dotés d'un GPS, où leur capacité de couverture est représentée par un modèle de couverture binaire. Les auteurs dans [11] ont proposé un algorithme de redistribution des capteurs, déployés initialement d'une manière aléatoire, basé sur la stratégie des forces virtuelles, pour atteindre la couverture globale du terrain tout en maintenant la connectivité du réseau. Pour ces mêmes objectifs, la solution de déploiement proposée dans [20] consiste à placer les capteurs à des positions situant sur une courbe spirale entourant la surface du terrain. L'inconvénient des approches de cette classe réside dans la modélisation très

simplifiée et irréaliste du terrain et la non prise en compte de son impact sur les performances du réseau.

2. APPROCHES DE DEPLOIEMENT BASEES SUR UN MODELE NUMERIQUE MATRICIEL DU TERRAIN

Afin de simplifier le recensement de solutions de déploiement de cette classe, nous les subdivisons en trois sous classes, selon le modèle de couverture adopté.

A. DEPLOIEMENT BASE SUR UN MODELE DE COUVERTURE BINAIRE

Les auteurs dans [19] ont utilisé l'algorithme de recuit simulé (SA) [23] pour maximiser le taux de couverture, où les capteurs sont supposés homogènes et omnidirectionnels, et leur capacité de détection est modélisée d'une manière binaire, en fonction de la distance et de la visibilité. Le recuit simulé a été conçu de deux manières : centralisée et distribuée. Dans la version centralisée, un schéma de déploiement est initialisé aléatoirement, et évolue en plusieurs itérations, par le déplacement d'un capteur quelconque à une position aléatoire sur le terrain pour améliorer la qualité de couverture. Dans la version distribuée, chaque nœud adopte à son niveau une solution initiale arbitraire et commence à générer des nouvelles solutions par le changement aléatoire de sa propre position en prenant en considération à chaque itération les nouvelles positions des voisins.

En utilisant les mêmes hypothèses et le même modèle de couverture adoptés dans [19], une solution de déploiement basé sur l'algorithme S-GA [28] a été proposée dans [60], visant à maximiser le taux de couverture, où un individu (chromosome), qui représente une solution valable au problème, est constitué de l'ensemble des positions occupées par les capteurs. Pour augmenter la probabilité d'améliorer la qualité de la solution pendant l'évolution de la population, les opérations de croisement ont été conçues de telle sorte que les positions des capteurs incluses dans les nouveaux chromosomes générés, soient bien réparties sur le terrain.

En supposant que les capteurs sont directionnels tout en adoptant un modèle de couverture binaire, formulé en fonction de la distance, de la visibilité et de l'orientation des capteurs, les auteurs dans [2] ont utilisé la méthode CMA-ES [21] pour déterminer les positions et les orientations des capteurs permettant de maximiser la couverture. Chaque individu, qui comporte les positions et les orientations des capteurs, est initialisé aléatoirement ; il commence à changer itérativement la position et/ou l'orientation d'un capteur sélectionné aléatoirement dans le but d'obtenir un schéma de déploiement plus efficace.

Dans [64], les auteurs ont proposé une extension de l'algorithme de déploiement proposé dans [69] pour les terrains 2D. Cet algorithme de base est basé sur le diagramme de Voronoï et fait recours au déplacement des capteurs pour éliminer les vides de couverture détectés. L'amélioration apportée consiste à prendre en considération les paramètres de la visibilité et d'obstacles pour rendre l'opération de déplacement des capteurs plus réaliste.

B. DEPLOIEMENT BASE SUR UN MODELE DE DETECTION HYBRIDE

Dans [5] [9] [15] [62], un modèle de couverture hybride, formulé en fonction de la distance et de la visibilité est adopté. Pour maximiser la qualité de couverture, les auteurs dans [5]

ont utilisé les algorithmes S-GA [28] et SS-GA [29], renforcés par la transformation 2D-DWT [37]. Dans ces deux méthodes, un individu qui représente une solution possible au problème est représenté par un vecteur qui contient les positions candidates des capteurs. La population des individus est initialisée aléatoirement et l'opération de croisements est effectuée en point unique. L'opérateur de mutation appliquée à un individu consiste à déplacer un capteur, sélectionné aléatoirement, vers une autre position dans la région la moins couverte, détectée par la transformation 2D-DWT.

Dans [9], les auteurs ont proposé l'algorithme CSO-WT basé sur la méthode d'optimisation CSO [54] renforcée par la transformation 2D-DWT [37] utilisée pour localiser et minimiser les vides de couverture et améliorer par conséquent la qualité de couverture au cours de l'évolution de la population.

Dans [15], les deux algorithmes ABC [35] et ABC-DSS [36] ont été utilisés pour maximiser la qualité de couverture, dans lesquels, la génération et l'évolution de la population des individus sont guidées par la méthode 2D-DWT [37] qui aide à localiser et éliminer les vides de couverture d'une façon similaire à celle utilisée dans CSO-DWT [9].

Dans [62], les auteurs ont proposé une solution de déploiement basée sur l'algorithme S-GA [28] renforcé par la transformation 2D-DWT [37] pour maximiser la qualité de couverture et la connectivité du réseau, évaluée en utilisant la méthode proposée dans [74]. La population des individus est initialisée d'une façon aléatoire. La sélection des individus est basée sur la couverture et la connectivité qu'ils produisent et les mutations sont guidées par la transformation 2D-DWT.

C. DEPLOIEMENT BASE SUR UN MODELE DE COUVERTURE PROBABILISTE

Sous l'hypothèse que les capteurs sont omnidirectionnels, et par l'adoption d'un modèle de couverture probabiliste formulé en fonction de la distance géodésique capteur-point, les auteurs dans [57] ont proposé un algorithme de déploiement glouton, qui se déroule en plusieurs itérations pour atteindre la k-couverture du terrain. À chaque itération, un capteur est placé à la position permettant de maximiser la somme de la qualité de couverture de tous les points du terrain. Si la qualité de couverture d'un point atteint un certain seuil, son niveau de couverture est incrémenté et sa qualité de couverture est remise à 0. Cette opération est répétée jusqu'à ce que tous les points aient un niveau de couverture prédéfini k ou bien tous les capteurs sont déployés.

Le travail réalisé dans [2] est amélioré dans [8] [14] par l'adoption d'un modèle de couverture probabiliste, formulé toujours en fonction de la distance, de la visibilité et de l'orientation des capteurs. Afin de maximiser la qualité de couverture globale, le recuit simulé SA [23] et CMA-ES [21] sont utilisées. La méthode SA démarre avec une solution initiale aléatoire, dénommée solution courante, comparée par la suite à des nouvelles solutions générées par le changement aléatoire de la position et de l'orientation d'un capteur quelconque. La solution courante v_{tmp} est mise à jour directement ou selon une probabilité, en fonction de sa qualité en comparaison à celle de la solution générée à chaque itération. Dans CMA-ES [21] un individu, qui représente une solution possible, comporte les positions et les orientations des capteurs. La population initiale des individus est générée aléatoirement et elle évolue en plusieurs itérations, où

l'individu qui maximise la qualité de couverture globale est sectionné comme une solution au problème.

Dans [13], la solution de déploiement proposée dans [2] est généralisée, en définissant les positions interdites où il est impossible de mettre les capteurs ainsi que l'ensemble des points critiques devant être couverts. Le modèle de couverture probabiliste proposé est conçu en fonction de la distance, de l'orientation du capteur et de la visibilité, évaluée cette fois d'une façon probabiliste en fonction du nombre d'obstacles. La méthode PSO [26], consolidée par la stratégie des forces virtuelles pour guider l'évolution de la population initiale des solutions générées aléatoirement, est utilisée pour trouver les positions et les orientations des capteurs, produisant un meilleur taux de couverture.

3. APPROCHES DE DEPLOIEMENT BASEES SUR UN MODELE NUMERIQUE TRIANGULAIRE DU TERRAIN

Dans [3] [4], une nouvelle solution de déploiement est proposée, visant à maximiser à la fois la qualité de couverture et la discrétion des capteurs tout en minimisant le coût de déploiement, sous l'hypothèse que les capteurs sont directionnels. Le modèle de couverture probabiliste utilisé est formulé en fonction de la distance, des caractéristiques du terrain, et de l'orientation du capteur. L'algorithme SS-GA [29] est utilisé afin de trouver le nombre minimum des capteurs, leurs positions et leurs orientations permettant de satisfaire les trois objectifs cités-supra, dont la population initiale, qui est générée aléatoirement et comporte des individus de tailles différentes, s'évolue selon des opérations de croisement et de mutation quasi-aléatoires.

Sous l'hypothèse que les capteurs sont homogènes et omnidirectionnels et par l'adoption d'un modèle de couverture binaire formulé en fonction de la distance, une solution de déploiement est proposée dans [7], visant à maximiser la qualité de couverture par un nombre défini des capteurs, en appliquant un partitionnement de Voronoï de la surface du terrain, où chaque capteur constitue le barycentre d'une cellule et assure tout seul sa couverture. Pour cela, la méthode bijective de Ricci [33] [34] est utilisée pour mapper la surface du terrain sur une surface plate. Par la suite, la partition de Voronoï de cette surface plate est calculée en utilisant des algorithmes conçus à cet effet [31] [32], pour déduire celle de la surface du terrain mappée.

En se basant sur les mêmes hypothèses adoptées dans [7], un algorithme de déploiement glouton est proposé dans [1] [10], où la surface du terrain est supposé partitionnée en morceaux de tailles relativement petites par rapport à la portée de couverture des capteurs. Son principe consiste à sélectionner les morceaux devant abriter les capteurs pour assurer la couverture totale du terrain. Pour cela, il est exécuté en plusieurs itérations ; dans chacune, un capteur est placé sur le morceau qui maximise la portion couverte du terrain.

Un tableau récapitulatif des principales approches de déploiement d'un RCSF sur un terrain tridimensionnel est joint au présent travail.

III. DESCRIPTION DE L'APPROCHE DE DEPLOIEMENT PROPOSEE

La principale contribution caractérisant la solution proposée consiste à analyser et partitionner le terrain, représenté par un

modèle numérique triangulaire, avant de procéder au déploiement des nœuds.

Nous utilisons trois types de nœuds, en l'occurrence, les capteurs, les relais et les collecteurs. Les capteurs sont les entités chargés de collecter les données et les acheminer vers le collecteur, en passant, si nécessaire, par les relais (passerelles), qui jouent le rôle de l'intermédiaire entre les capteurs et les collecteurs. Les capteurs sont supposés homogènes et omnidirectionnels, où leur capacité de couverture est formulée par un modèle binaire.

Les relais, dédiés uniquement pour renforcer la connectivité du réseau, sont supposés homogènes et disposent d'une capacité de communication nettement supérieure à celle des capteurs. On suppose également que les collecteurs prennent des positions arbitraires sur le terrain.

L'objectif consiste à trouver les meilleures positions à occuper par les capteurs et les relais afin de maximiser la couverture et la connectivité en prenant en considération la topographie réelle du terrain. Deux phases indépendantes sont donc envisagées, la première concerne le déploiement des capteurs et la deuxième concerne le déploiement des relais.

Ces deux phases de déploiement sont précédées et guidées par une opération de partitionnement du terrain fourni sous forme d'un modèle triangulaire. A cet effet, une heuristique de partitionnement est proposée, consiste à procéder à la fusion des triangles, en analysant la visibilité inter-triangles, ce qui donne à la fin des régions ayant des topographies relativement simples.

Par la suite, la 1^{ère} phase est entamée, en procédant au déploiement des capteurs dans chaque zone, d'une façon indépendante, selon deux algorithmes de déploiement conçus à cet effet à savoir : une méthode de distribution uniforme et le recuit simulé, dans l'objectif de maximiser le taux de couverture.

La 2^{ème} phase qui concerne le déploiement des relais vise à renforcer davantage la connectivité du réseau. Les relais doivent occuper des positions ayant une large visibilité, situant sur les lignes frontalières (lignes de crêtes) séparant les régions construites. Les détails de cette 2^{ème} phase et son évaluation feront l'objet d'un autre travail, à réaliser au cours de la 3^{ème} année.

Quelques résultats portés sur la figure jointe au présent travail démontre la pertinence de la méthode de partitionnement qui précédé le déploiement des capteurs, où elle a permis d'améliorer davantage le taux de couverture produit par la méthode de distribution uniforme et le recuit simulé.

IV. CONCLUSION ET TRAVAUX FUTURS

Le recensement des solutions de déploiement d'un réseau de capteurs sans fil pour des environnements réalistes nous a permis de souligner quelques lacunes, notamment dans la manière de formuler l'impact de la topographie sur la couverture et la connectivité.

A cet effet, nous avons proposé une nouvelle approche, qui consiste à analyser en premier lieu le terrain et le partitionner selon des critères de visibilité, afin d'alléger l'impact de la topographie et augmenter ainsi la couverture de la zone d'intérêt. La connectivité sera prise en charge dans le prochaine travail qui consiste à exploiter la phase de partitionner pour sélectionner les meilleures positions devant

abriter les relais, qui assurer ainsi la communication inter-régions.

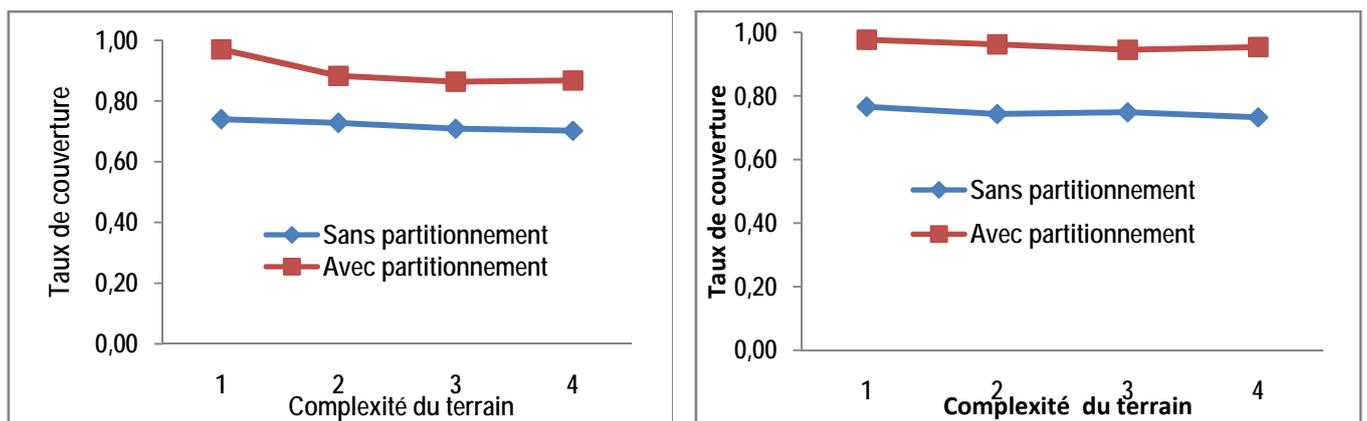
REFERENCES

- [1] Zhao, M.C., Lei, J., Wu, M.Y., Liu, Y., Shu, W. : "Surface coverage in wireless sensor networks". Proc. of IEEE INFOCOM 2009, 2009, pp. 09–117;
- [2] V. Akbarzadeh, A. Ko, C. Gagné, and M. Parizeau, "Topography-aware sensor deployment optimization with CMA-ES," in Proc. 11th PPSN, 2010, pp. 141–150.
- [3] H. Topcuoglu, M. Ermis, and M. Sifyan, "Positioning and utilizing sensors on a 3D terrain Part I - Theory and modeling," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 41, no. 3, pp. 376–382, May 2011;
- [4] H. Topcuoglu, M. Ermis, and M. Sifyan, "Positioning and utilizing sensors on a 3D terrain: Part II - Solving with a hybrid evolutionary algorithm," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., 2010;
- [5] N. Unaldi, S. Temel, V.K. Asari, "Method for Optimal Sensor Deployment on 3D Terrains Utilizing a Steady State Genetic Algorithm with a Guided Walk Mutation Operator Based on the Wavelet Transform", Sensors, 12(4), 5116-5133, 2012;
- [6] L. Liu and H. Ma, "On Coverage of Wireless Sensor Networks for Rolling Terrains," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 1, pp. 118-125, Jan. 2012;
- [7] M. Jin, G. Rong, H. Wu, L. Shuai, and X. Guo, "Optimal Surface Deployment Problem in Wireless Sensor Networks," Proc. IEEE INFOCOM, pp. 2345-2353, 2012;
- [8] Akbarzadeh V., Gagne C., Parizeau M., Argany M., Mostafavi M. A., "Probabilistic Sensing Model for Sensor Placement Optimization Based on Line-of-Sight Coverage", IEEE Transactions on Instrumentation and Measurement, vol. 62, num. 2, p. 293-303, February 2013;
- [9] Temel S., Unaldi N., Kaynak O., "On Deployment of Wireless Sensors on 3D Terrains to Maximize Sensing Coverage by Utilizing Cat Swarm Optimization with Wavelet Transform", IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, num. 1, p. 111-120, January 2014;
- [10] L. Kong, Y.Liu, W.Shu, "Surface coverage in sensor networks," IEEE Trans.Parallel Distrib. Syst., vol. 25, no. 1, pp. 234–243, Jan. 2014;
- [11] Nadia Boufares, Ines Khoufi, Pascale Minet, Leila Saidane. "3D Surface Covering with Virtual Forces". Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN), Nov 2015, Hammamet, Tunisia;
- [12] F. Li, J. Luo, W. Wang, and Y. He, "Autonomous Deployment for Load Balancing k-Surface Coverage in Sensor Networks," Wireless Communications, IEEE Transactions on, vol. 14, no. 1, pp. 279–293, 2015;
- [13] "Optimization for the sensor placement problem in 3D environments", 2015;
- [14] V. Akbarzadeh, C. Gagné, M. Parizeau, and M. Mostafavi, "Black-box optimization of sensor placement with elevation maps and probabilistic sensing models," in Proc. Int. Symp. ROSE, 2011, pp. 89–94;
- [15] "A Novel sensor deployment method based on image processing and Wavelet Transform to optimize the surface coverage in WSNs", 2016;
- [16] "Effect of 3D Topographical Surfaces for the Performance Evaluation of Wireless Sensor Networks", 2008;
- [17] Sink placement on a 3D terrain for border surveillance in wireless sensor networks, 2012 ;
- [18] S. Oktug, A. Khalilov, and H. Tezcan, "3D Coverage Analysis under Heterogeneous Deployment Strategies in Wireless Sensor Networks," Proc. Fourth Advanced Int'l Conf. Telecomm. (AICT '08), 2008;
- [19] "Guiding sensor-node deployment over 2.5D Terrain", 2015;
- [20] "Mountainous terrain coverage in mobile sensor networks", 2015;
- [21] N. Hansen and A. Ostermeier. "Completely derandomized self-adaptation in evolution strategies". Evolutionary Computation, 9(2):159-195, 2001;
- [22] J. Nocedal and S. Wright, "Numerical optimization". Springer verlag, 1999;
- [23] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," science, vol. 220, no. 4598, p. 671, 1983;

- [24] M. Hefeeda and H. Ahmadi, "Energy efficient protocol for deterministic and probabilistic coverage in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 99, pp. 579–593, 2010;
- [25] X. Bai, S. Kumar, D. Xuan, Z. Yun, and T. Lai, "Deploying wireless sensors to achieve both coverage and connectivity," in *Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2006, pp. 131–142;
- [26] Kennedy, J. and R. Eberhart, "Particle swarm optimization," *Proc. of the IEEE International Conference on Neural Networks*, pp.601-610, 1995;
- [27] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995;
- [28] Cobb, H.; Grefenstette, J. "GA for tracking changing environments". In *Proceedings of the 5th International Conference on GA*; Morgan Kaufman: San Francisco, CA, USA, 1993;
- [29] Whitley, D.; Kauth, J. "GENITOR: A different genetic algorithm". In *Proceedings of the Rocky Mountain Conference on Artificial Intelligence*, Denver, CO, USA, 1988;
- [30] L. Chen and M. Holst, "Efficient mesh optimization schemes based on optimal Delaunay Triangulations", *Computer Methods in Applied Mechanics and Engineering* 200(9) 967-984, 2011;
- [31] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982;
- [32] Q. Du, M. Emelianenko, and L. Ju, "Convergence of the Lloyd Algorithm for Computing Centroidal Voronoi Tessellations," *SIAM J. Numer. Anal.*, vol. 44, no. 1, pp. 102–119, 2006.
- [33] B. Chow and F. Luo, "Combinatorial Ricci Flows on Surfaces," *Journal Differential Geometry*, vol. 63, no. 1, pp. 97–129, 2003.
- [34] R. S. Hamilton, "Three manifolds with positive Ricci curvature," *Journal of Differential Geometry*, vol. 17, pp. 255–306, 1982.
- [35] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm", *Journal of Global Optimization*, Vol.39, No.3, pp.459–471, 2007;
- [36] D. Karaboga and B. Akay, "A modified Artificial bee colony (ABC) algorithm for constrained optimization problems", *Applied Soft Computing*, Vol.11, No.3, pp.3021–3031, 2011;
- [37] R.J. Colom-Palero, R. Gadea-Girones, F.J. Ballester-Merelo and M. Martinez-Peiro, "Flexible architecture for the implementation of the two-dimensional discrete wavelet transform (2D-DWT) oriented to FPGA devices", *Microprocessors and Microsystems*, Vol.28, No.9, pp.509–518, 2004.
- [38] "<http://srtm.csi.cgiar.org/Index.asp>", 2011;
- [39] N. George and W. Sharad, "Geographic Data Processing," *Computing Surveys*, vol. 11, no. 2 pp. 139-181, 1979;
- [40] T. K. Peucker and D. H. Douglas, "Detection of surface-specific points by local parallel processing of discrete terrain elevation data". *Computer Graphics and Image Processing*, 4:375–387, 1975;
- [41] T. K. Peucker, R. J. Fowler, J. J. Little, and D. M. Mark. "The triangulated irregular network". In *Proceedings, American Society of Photogrammetry: Digital Terrain Models (DTM) Symposium*, pages 516–540. American Society for Photogrammetry & Remote Sensing, May 1978;
- [42] Shuttle Radar Topography Mission (SRTM). <http://www2.jpl.nasa.gov/srtm/>;
- [43] Light Detection and Ranging data (LIDAR). <http://atlas.lsu.edu/lidar/>;
- [44] L.A. Santalo, "Integral Geometry and Geometric Probability". Addison-Wesley, 1976;
- [45] D.S. Hochbaum and W. Maass, "Approximation Schemes for Covering and Packing Problems in Image Processing and Vlsi," *J. ACM*, vol. 32, no. 1, pp. 130-136, 1985;
- [46] M. D. Proctor and W. J. Gerber, "Line-of-sight attributes for a generalized application program interface," *J. Defense Model. Simul.*, vol. 1, no. 1, pp. 43–57, 2004
- [47] D. Handerson, "Modterrain: A proposed standard for terrain representation in entity level simulation," Master's thesis, Naval Postgraduate School, Monterey, CA, 1999;
- [48] L. D. Floriani and P. Magillo. "Algorithms for visibility computation on digital terrain models". In *SAC*, pages 380–387, 1993;
- [49] J. Wang, G. J. Robinson, and K. White. "Generating viewsheds without using sightlines". *Photogrammetric Engineering & Remote Sensing*, 66(1):87–90, January 2000;
- [50] Moller, T. and B. Trumbore, "Fast, Minimum Storage Ray/Triangle Intersection", *Journal of Graphics Tools*, Vol. 2, No.1, pp.21-28, 1997;
- [51] X.W. Liu and K. Cheng, "Three-dimensional extension of Bresenham's algorithm and its application in straight-line interpolation", *Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture*, Vol.216, No.3, pp.459–463, 2002;
- [52] Wang, J., G.J. Robinson, and K. White, 1996. "A fast solution to local viewshed computation using grid-based digital elevation models", *Photogrammetric Engineering & Remote Sensing*, 62:1157-1164.
- [53] Leila De Floriani, Paola Marzano & Enrico Puppo, "Line-of-sight communication on terrain models", *International Journal of Geographical Information Systems*, 8:4, 329-342, 1994.
- [54] S. C. Chu and P. W. Tsai, "Computational intelligence based on the behaviour of cats," *Int. J. Innovative Comput. Inform. Contr.*, vol. 3, no. 1, pp. 163–173, Feb. 2007;
- [55] C. H. Wu, K. C. Lee, and Y. C. Chung, "A Delaunay triangulation based method for wireless sensor network deployment," *Comput. Commun.*, vol. 30, no. 14, pp. 2744–2752, Oct. 2007;
- [56] Lassila, P., Hyttiä, E., Koskinen, H.: "Connectivity properties of random waypoint mobility model for ad hoc networks". In *MedHoc-Net*, Ile de Porquerolles, 2005, pp. 159–168.
- [57] Song, T., Gong, C., Liu, C., 2012. "A practical coverage algorithm for wireless sensor networks in real terrain surface". *Int. J. Wireless Mobile Comput.* 5 (4), 358–364;
- [58] "SURF: A Connectivity-based Space Filling Curve Construction Algorithm in High Genus 3D Surface WSNs", 2015;
- [59] "Surface coverage algorithm in directional sensor networks for three-dimensional complex terrains", 2016;
- [60] "Process of 3D wireless decentralized sensor deployment using parsing crossover scheme", 2015;
- [61] "An Efficient Large-Scale Sensor Deployment Using a Parallel Genetic Algorithm Based on CUDA", 2015;
- [62] "Wireless Sensor Deployment Method on 3D Environments to Maximize Quality of Coverage and Quality of Network Connectivity", 2014;
- [63] "Design of a WSN for the Sampling of Environmental Variability in Complex Terrain", 2014;
- [64] "3D extension of the VOR algorithm to determine and optimize the coverage of geosensor networks", 2014;
- [65] "Wireless Sensor Deployment using Modified Discrete Binary PSO Method", 2013;
- [66] "Mobility increases the surface coverage of distributed sensor networks", 2013;
- [67] "Genetic Algorithm-Based 3D Coverage Research in Wireless Sensor Networks", 2013;
- [68] "A Coverage Strategy for Wireless Sensor Networks in a Three-dimensional Environment", 2013;
- [69] Wang, G., Cao, G., La Porta, T. "Movement-Assisted Sensor Deployment". *Mobile Computing*, *IEEE Transactions* 5, 640 – 652, 2006;
- [70] "Adaptive PSO based Algorithm for Optimal WSN Deployment in 3 Dimensional Terrains", 2012;
- [71] "A Context-Aware Method for Sensor Network Deployment and Coverage Optimization", 2012;
- [72] "Multiresponse Optimization of Stochastic WSN Deployment Using Response Surface Methodology and Desirability Functions", 2010;
- [73] "Efficient Power Conservation Mechanism in Spline Function Defined WSN Terrain", 2014;
- [74] De Marco, G., "MOGAMESH: A multi-objective algorithm for node placement in wireless mesh networks based on genetic algorithms," *Wireless Communication Systems*, 2009. *ISWCS 2009. 6th International Symposium on*, vol., no., pp.388-392, 7-10 Sept. 2009;
- [13] Elfes, A., "Occupancy Grids: A Stochastic Spatial Representation for Active Robot Perception", *Autonomous Mobile Robots: Perception, Mapping and Navigation*, 1, 60-70, 1991.

Réf.	Modèle terrain	Modèle couverture	Type capteurs	Objectifs	Méthodes résolution
[11]	Mathématique	Binaire	Omnidirectionnel	Couverture, Connectivité	Forces virtuelles
[20]	Mathématique	Binaire	Omnidirectionnel	Couverture, Connectivité	Modèle mathématique
[19]	Numérique matriciel	Binaire	Omnidirectionnel	Couverture	Recuit simulé (SA)
[60]	Numérique matriciel	Binaire	Omnidirectionnel	Couverture	S-GA
[2]	Numérique matriciel	Binaire	Directionnel	Couverture	CMA-ES
[64]	Numérique matriciel	Binaire	Omnidirectionnel	Couverture, connectivité	Algorithme de Voronoï
[5]	Numérique matriciel	Hybride	Omnidirectionnel	Couverture	S-GA, SS-GA
[9]	Numérique matriciel	Hybride	Omnidirectionnel	Couverture	CSO
[15]	Numérique matriciel	Hybride	Omnidirectionnel	Couverture	ABC, ABC-DSS
[62]	Numérique matriciel	Hybride	Omnidirectionnel	Couverture, connectivité	S-GA
[57]	Numérique matriciel	Probabiliste	Omnidirectionnel	Couverture	Algorithme glouton
[8]	Numérique matriciel	Probabiliste	Directionnel	Couverture	SA, CMA-ES
[14]	Numérique matriciel	Probabiliste	Directionnel	Couverture	SA, CMA-ES
[13]	Numérique matriciel	Probabiliste	Directionnel	Couverture	PSO
[3]	Numérique triangulaire	Probabiliste	Directionnel	Couverture, coût, discrétion des capteurs	SS-GA
[4]	Numérique triangulaire	Probabiliste	Directionnel	Couverture, coût, discrétion des capteurs	SS-GA
[7]	Numérique triangulaire	Binaire	Omnidirectionnel	Couverture	Algorithme de Voronoï
[1]	Numérique triangulaire	Binaire	Omnidirectionnel	Couverture	Algorithme glouton
[10]	Numérique triangulaire	Binaire	Omnidirectionnel	Couverture	Algorithme glouton

Tableau 1. Approches de déploiement d'un RCSF sur un terrain tridimensionnel (récapitulatif).



A. Distribution uniforme.

B. Recuit simulé.

Figure 1. Quelques résultats obtenus, montrant que la couverture est améliorée davantage par la phase de partitionnement, pour les deux méthodes de résolution utilisées.

Approche générale pour la détection des cyber-attaques

Mohamed Amine DAOUD

Directeurs de thèse : Dr. Youcef DAHMANI

Résumé : un des problèmes mondiaux dans le domaine de l'informatique est bien la sécurité de l'information. Le développement des techniques et des politiques est devenu un outil incontournable pour la sécurité des systèmes contre les attaques. Ce travail s'inscrit dans une optique d'application de la sécurité des systèmes informatiques. Dans ce contexte, le souhait est de réaliser une solution capable de gérer la surveillance et la protection universel. Les différentes solutions doivent être combinées et être améliorées pour être efficace contre une telle attaque pour sécuriser un tel service. Notre amélioration repose sur l'intégration d'une de ces techniques d'apprentissage artificiel de machine tels que les réseaux bayésiens. La rapidité et la performance sont des facteurs très importants pour la détection des attaques et l'amélioration de la qualité des logiciels, tout ça pour augmenter le niveau de la cybersécurité.

I. INTRODUCTION

Ces dernières années et même actuellement, les cyberattaques sont devenues de plus en plus complexes, sophistiquées et difficiles à détecter. Elles menacent les processus industriels, les infrastructures critiques, les entreprises et les individus, elles peuvent entraîner des graves pertes de données économiques et surtout sur la sécurité des employés et des citoyens. L'interconnexion des systèmes informatiques, par les réseaux, donne la possibilité à des nouvelles sources des cyber-menaces et ouvre des nouvelles voies pour mener des attaques contre des destinations tel que l'Iran sur son programme nucléaire qui a été interrompu par un code cybernétique malveillant ou bien la Chine et la Russie qui ont été soupçonnées de voler des données militaires secrètes à partir des ordinateurs des états unies. Ces événements se sont produits parce que le cyberspace offre un moyen rapide et peu coûteux d'utiliser une cyber-attaque comme des outils pour exploiter les faiblesses significatives, ces outils peuvent permettre à l'espionnage de modifier l'équilibre du pouvoir économique ou militaire, de contrôler et endommager secrètement les équipements. Une fois, les chercheurs ont vu que le taux des cyber-attaques augmente en quantités et en qualités sur les systèmes informatiques et les réseaux, ils n'ont pas arrêté de produire et développer des idées et de rechercher des solutions pour contrer ces attaques, dans ce cas, un nouveau terme a été inventé ce qu'on appelle la détection.

II. PROBLEMATIQUE

L'insuffisance des systèmes traditionnels de détection des intrusions à détecter de nouveaux types des cyber-attaques qui n'existent pas dans leurs bases de données et en plus les lacunes des systèmes de détection qui donnent une possibilité d'améliorer ces systèmes pour fournir un niveau maximum de protection contre ces cyber-attaques. Notre objectif décrit une démarche qui permet d'appréhender et de construire une politique de sécurité et de choisir la solution adéquate à ces besoins de sécurité. Notre amélioration repose sur l'intégration d'une de ces techniques d'apprentissage artificiel de machine tels que les réseaux bayésiens, les réseaux de neurones ou les chaînes de Markov etc à ces systèmes au fil du temps, soit sur la base de leur propre expérience, soit à partir de données antérieures fournies par d'autres programmes et de fournir un outil qui va permettre de soutenir les parties prenantes tout au long du processus. Des systèmes robustes avec une rigueur scientifique et technique des technologies de détection permettraient de mieux comprendre les caractéristiques du modèle. Cette connaissance à son tour pourrait permettre de construire des modèles mathématiques, statistiques ou probabilistiques.

Dans une approche de détection des attaques, il faudrait prendre des mesures efficaces pour réduire les risques et les dangers au fil du temps. Notre travail est d'améliorer la détection des cyber-attaques afin de rechercher des nouvelles attaques probables. Ainsi, dans le cadre d'une première étape vers l'élaboration de mesures de sécurité des systèmes, il faut comprendre les impacts de ces cyber-attaques et les pannes du système qui en résultent à cause de ces cyber-attaques. Pour assurer une sécurité d'un système informatique, il est constaté de contrer les attaques, qui étaient jusqu'à présent presque exclusivement logicielles, il faut tenir compte d'une bonne connaissance de la plateforme matérielle qui devient de plus en plus importante pour une meilleure maîtrise de la sécurité d'un système informatique

III. TRAVAUX CONNEXES

Durant notre première année, nous avons abordé des travaux existants récents sur la détection des cyber-attaques. Ces dernières se focalisent sur plusieurs domaines car ils sont mal protégés ou mal maintenus. Parmi ces domaines, on cite les infrastructures critiques [1],[2],[3],[4],[5], le Cloud Computing [6],[7],[8],[9],[10] Smart Home [11],[12],[13] et les réseaux sans fil [14]. Ces domaines ont pris une grande importance par les chercheurs pour la proposition des approches, des méthodes et des solutions pour ces domaines

pour les renforcer du côté de la sécurité. Nous avons synthétisé les avantages de chaque domaine et on a présenté les problèmes dont souffre la cybersécurité. Nous avons vu aussi un profil de ces travaux sur les méthodes et les approches proposées pour aider à augmenter le taux de la cybersécurité chacun dans son domaine. Il s'agira de recueillir les besoins et de comprendre les problèmes de décision liés à une situation de cyber-attaque sur un système.

IV. PERSPECTIVES

Notre travail de recherche se focalisera sur deux axes très intéressants.

Premièrement, un état de l'art sur les cyber-attaques sur le domaine des réseaux choisis. La connaissance à priori des cyber-attaques, ou bien savoir le fonctionnement de l'attaque, l'identification des menaces et des vulnérabilités sont nécessaires pour la sécurité des systèmes. Elles sont difficiles à caractériser car elles ont un degré élevé d'indétermination et ne peuvent pas être évitées. Un élément clé d'une défense devrait inclure des renseignements sur les menaces, pour partager et recueillir des renseignements sur les nouvelles menaces et les menaces émergentes ciblant la sécurité des réseaux.

Deuxièmement, proposer une solution de l'approche de modélisation des cyber-attaques. Les vulnérabilités sont des brèches qui peuvent être détectées et contrôlées. Chaque attaquant a un ensemble d'attributs tels que la connaissance du système cible, les ressources techniques, la motivation pour mener une attaque contre un système pour sécuriser un tel système informatique ou un réseau, il faut tout d'abord comprendre le mouvement ou bien le profil d'un attaquant. Ce travail peut distinguer les différents types de cyber-attaques, ceci se rapportent aux processus cognitifs et leurs implications pour améliorer la cybersécurité. Cependant, les connaissances ont des contributions lorsqu'on juge si une séquence d'événements représente une cyber-attaque [15], [16], [17], [13].

V. CONCLUSION

À travers ce rapport, nous avons présenté d'état d'avancement de notre objectif. Nous avons présenté nos problématiques de notre sujet et nous avons abordé les travaux existants récents de la détection des cyber-attaques. Ainsi, on a fixé comme perspectives le but de la prochaine année.

Références

- [1] Y. Nong, T. Farley, "A Scientific Approach to Cyberattack Detection", Arizona State University, IEEE November 2005.
- [2] S. Biswas and A. Sarwat, "Vulnerabilities in Two-Area Automatic Generation Control Systems under Cyberattack", ISBN 978-1-5090-2002-7, IEEE, pp 40-46, 2016
- [3] P. Nader, P. Honeine et P. Beauseroy, "Detection of Cyberattacks In a Water Distribution System Using Machine Learning Techniques", ISBN: 978-1-4673-7504-7, IEEE, 2016.
- [4] E. Pricop, S. F. Mihalache, "Fuzzy approach on modelling cyber-attacks patterns on data transfer in industrial control systems",

- ECAI 2015, International Conference – 7th Edition Electronics, Computers and Artificial Intelligence, IEEE, 2015
- [5] T. Morita, S. Yogo, M. Koike, T. Hamaguchi, S. Jung, I. Koshijima, Y. Hashimoto, "Detection of Cyber-Attacks with Zone Dividing and PCA", 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013, Elsevier, pp 727 – 736, 2013
- [6] H. Badis, G. Doyen and R. Khatoun, "A collaborative approach for a source based detection of botclouds," IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015, pp. 906-909.
- [7] R. Bose, X. Luo, Y. Liu, "The Roles of Security and Trust: Comparing Cloud Computing and Banking", The 2nd International Conference on Integrated Information, Elsevier, pp 30-34, 2013.
- [8] S. Bologna, A. Fasani and M. Martellini, "Cyber Security and Resilience of Industrial Control Systems and Critical Infrastructures", Cyber Security, Springer, 2013, pp 57-72.
- [9] R. Velumadhava Rao, K. Selvamani, "Data Security Challenges and Its Solutions in Cloud Computing", (ICCC-2014), Elsevier, pp 204-209, India
- [10] M. Abdhamed, K. Kifayat, Q. Shi, W. Hurst, "A System for Intrusion Prediction in Cloud Computing", ICC '16, ACM, 2016, Cambridge, United Kingdom, 2016
- [11] Y. Liu, S. Hu, T. Ho, "Vulnerability Assessment and Defense Technology for Smart Home Cybersecurity Considering Pricing Cyberattacks", IEEE, 2014
- [12] Y. Liu, S. Hu, A. Y. Zomaya, "The Hierarchical Smart Home Cyberattack Detection Considering Power Overloading and Frequency Disturbance", IEEE, 2016.
- [13] M. Miller, D. Haglund, G. Lee, "Smart grid cyber security", ISGAN, 2012.
- [14] M. N. Mohammed, N. Sulaiman, "Intrusion Detection System Based on SVM for WLAN", Procedia Technology, Elsevier, 2012.
- [15] M. A. Daoud and Y. Dahmani, "ARTT taxonomy and cyber-attack Framework," First International Conference on New Technologies of Information and Communication (NTIC), Mila, 2015, pp. 1-6. NTIC.2015.7368742, IEEE, 2015.
- [16] N. Ben-Asher, C. Gonzalez, "Effects of cyber security knowledge on attack detection", Computers in Human Behavior 48, Elsevier, pp 51–61, 2015.
- [17] H. Buijn, M. Janssen, "Building cybersecurity awareness: The need for evidence-based framing strategies", Government Information Quarterly, Elsevier, pp 1-7, 2017.
- [18] Cybercrime and Business., "The Global Cybercrime Industry", CHAPTER 1, Elsevier, 2017

Système crypto-biométrique pour la sécurité de l'information

Othaila CHERGUI

Directeurs de thèse : Hakim BENDJENNA

Équipe de Recherche: Systèmes Intelligents et Distribués (SID)

Résumé— La sécurité des informations constitue une priorité majeure. La combinaison de la cryptographie et la biométrie fournit la liaison de caractéristiques personnelles de l'utilisateur à des clés cryptographiques. Dans cet article, nous allons présenter les deux projets qui sont réalisés au cours de la 2^{ème} année (système crypto-biométrique appliqué à l'ATM et un système de tatouage numérique).

I. INTRODUCTION

La cryptographie et la biométrie sont préconisées comme des solutions parmi les plus intéressantes pour assurer la sécurité et considérées comme l'un des moyens les plus sûrs.

D'une part, la cryptographie est un ensemble de méthodes et techniques mathématiques permettant de sécuriser des messages. L'idée principale de la cryptographie est de coder le message de façon à ce qu'il devienne secret en utilisant une clé (le chiffrement). L'opération inverse (le déchiffrement) nécessite quant à elle une clé pour obtenir le message original. Sans connaissance préalable de la clé, le déchiffrement sera impossible [1].

D'autre part, la biométrie est un ensemble de technologies qui exploitent des caractéristiques humaines physiques, biologiques ou comportementales pour différencier des personnes [2]. Elle permet la reconnaissance de l'identité d'une personne dans le but de contrôler l'accès à une information. Contrairement aux moyens traditionnels qui sont basés sur des connaissances ou des possessions, la biométrie est basée sur des caractéristiques liées à la personne elle-même. Elle constitue un lien fort et permanent entre une personne et son identité ; elle ne pourrait pas être volée, oubliée, perdue ou transmise à une autre personne. C'est la raison pour laquelle elle est considérée comme la manière la plus sécurisée par rapport à l'utilisation des moyens classiques tels que les cartes ou les mots de passe.

En conclusion, la cryptographie et la biométrie sont deux moyens puissants pour assurer la sécurité de l'information. Les deux moyens peuvent être combinés afin d'être complémentaires. Cette combinaison a principalement deux objectifs [3] à savoir : la sécurisation des clés cryptographiques et la protection du template biométrique. Pour notre travail, nous nous intéressons à la sécurisation de la clé au sein du réseau en utilisant les techniques biométriques.

II. PROBLÉMATIQUE

Le principe de la cryptographie est basé sur une clé de chiffrement. La transformation inverse n'est possible qu'avec la connaissance de la clé de déchiffrement. Donc, la sécurité dans les systèmes cryptographiques dépend de la sécurité de cette clé. Les clés courtes peuvent être facilement devinées ou

piratées, ce qui exige l'utilisation de clés longues et aléatoires pour atteindre une haute sécurité. Ceci les rend difficiles à mémoriser par les utilisateurs. De ce fait, les clés doivent être sauvegardées sur un support physique (carte à puces) [4]. En outre, certains mécanismes d'authentification sont utilisés afin de limiter l'accès à ces clés. Les mécanismes traditionnels d'authentification utilisés en cryptographie sont basés sur une connaissance ou une possession. Ils sont affectés à l'identité de l'utilisateur sans qu'il soit nécessairement présent. Par conséquent, le système ne peut pas faire la distinction entre l'attaquant et un utilisateur légitime [5]. Donc, assurer la sécurité de la clé cryptographique reste un défi principal des systèmes cryptographiques (le cas du chiffrement symétrique).

Les caractéristiques biométriques sont uniques pour chaque individu et sont strictement liées à l'utilisateur [6], d'où l'utilisation de la biométrie afin de contrôler l'accès aux clés. Cette dernière offre une solution fiable au problème d'authentification dans les systèmes cryptographiques et la sécurité de ces derniers serait ainsi prouvée [7].

III. TRAVAUX RÉALISÉS

Grâce à l'avancée technologique qui a touché de nombreux domaines, la sécurité d'information est devenue une exigence de plus en plus répandue. Dans cet article, nous présenterons deux systèmes : le 1^{er} vise à améliorer la sécurité des transactions bancaires électroniques, tandis que le 2^{ème} s'intéresse aux droits d'auteur des images (tatouage numérique).

A. L'amélioration de la sécurité des transactions bancaires électroniques

L'ATM (Automated teller machine) est une technologie bancaire qui offre un service 24/24 heures et des transactions simples. Cependant, l'ATM connecte le client avec son compte bancaire en accédant aux comptes d'un système bancaire, ce qui exige que ce dernier soit sécurisé. Le système d'ATM existant utilise une carte associée à un code PIN. Ainsi, ces deux méthodes traditionnelles d'authentification risquent d'être volées, perdues, oubliées ou piratées [8].

Notre système crypto-biométrique proposé [9] relie l'utilisateur à son identité, ce qui présente un dépassement des limites des méthodes traditionnelles. Par ailleurs, la cryptographie est ajoutée afin de renforcer l'authentification et de rendre la transaction ATM plus sûre. En effet, nous proposons ce système ATM afin d'assurer la sécurité des informations bancaires et des transactions lorsque l'ATM est liée au réseau public.

Dans ce travail, nous avons opté pour l'empreinte palmaire comme modalité utilisée pour l'authentification biométrique. D'ailleurs, nous extrayons les caractéristiques de l'empreinte

palmaire en utilisant la transformée en cosinus discrète (DCT) qui permet d'exploiter efficacement le contenu des images. Ensuite, la quantification vectorielle est appliquée sur les vecteurs caractéristiques afin de les représenter sous forme d'ensemble de vecteurs de référence (codeword). Notre système ATM proposé ajoute au système ATM existant l'authentification biométrique et encode les transactions par la cryptographie (Figure 1).



Fig 1. Le système bancaire proposé.

Ce système fonctionne en deux phases : phase d'enrôlement et phase d'authentification. L'enrôlement est la phase où les clients de la banque sont enregistrés dans le système pour la 1^{ère} fois. Le processus d'enrôlement est décrit dans la Figure 2.

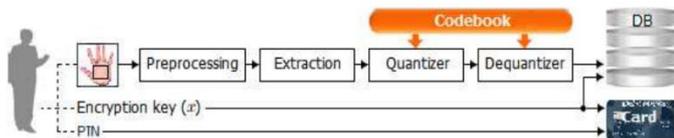


Fig 2. La phase d'enrôlement.

La phase d'authentification correspond à la phase opérationnelle du système ATM proposé. Dans cette phase, les clients de la banque passent leurs possibles transactions.

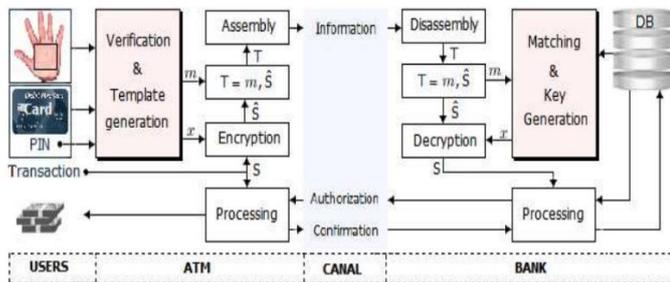


Fig 3. Authentification des clients.

Le système proposé est évalué et testé en utilisant une base de données de 250 personnes. Les résultats prouvent un niveau de sécurité élevé du système proposé (Figure 4).

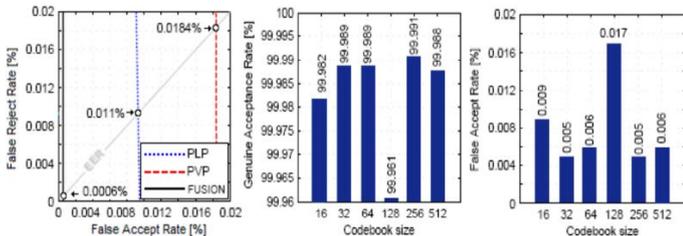


Fig 4. Quelques résultats du test du système proposé.

B. Le tatouage numérique

Le 2^{ème} système proposé s'intéresse aux droits d'auteur d'images. En effet, l'empreinte palmaire du propriétaire de l'image est utilisée pour construire une marque qui est insérée dans l'image de façon invisible et de telle sorte que la marque n'influe pas sur la qualité d'image. Plusieurs méthodes sont

utilisées et testées afin de sélectionner la meilleure. Ce système comporte deux phases : phase d'insertion de la marque et phase de vérification du propriétaire de l'image (Figure 5 et Figure 6).

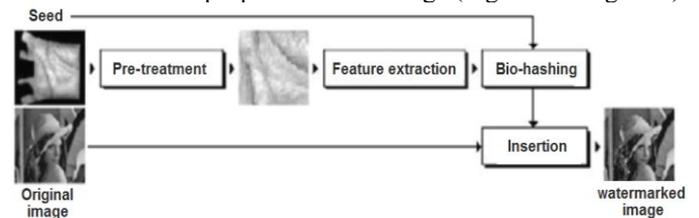


Fig 5. Processus d'insertion de la marque.

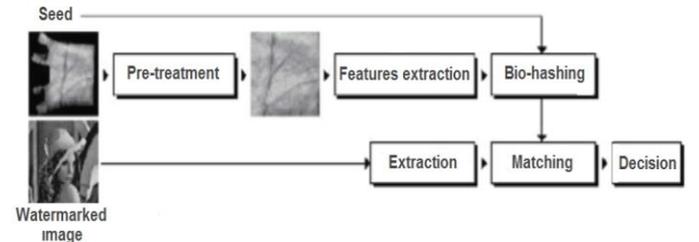


Fig 6. Processus de vérification de la marque.

IV. PERSPECTIVES

Les prochains travaux à réaliser sont les suivants :

- Publier le système ATM proposé dans une revue.
- Réaliser un système crypto-biométrique qui appartient au mode key-binding.
- Utiliser d'autres méthodes afin d'améliorer le système proposé.

Références

- [1] Kanade, Sanjay G., Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. "Enhancing information security and privacy by combining biometrics with cryptography." *Synthesis Lectures on Information Security, Privacy, and Trust 3.1* (2012): 1-140.
- [2] L. Wang, "Some issues of biometrics: technology intelligence, progress and challenges", *International Journal of Information Technology and Management*, Inderscience publisher, Vol. 11. 1/2, pp. 72-72, 2012.
- [3] Maiorana Emanuele. "Biometric cryptosystem using function based on-line signature recognition". *Expert Systems with Applications*, vol. 37, no 4, pp. 3454-3461, 2010.
- [4] Boddeti, Vishnu Naresh, Fei Su, and BVK Vijaya Kumar. "A biometric key-binding and template protection framework using correlation filters." *International Conference on Biometrics*. Springer Berlin Heidelberg, 2009.
- [5] Kanade, Sanjay, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. "Multi-biometrics based cryptographic key regeneration scheme." *Biometrics: Theory, Applications, and Systems*, 2009. BTAS'09. IEEE 3rd International Conference on. IEEE, 2009.
- [6] Chae, Seung-Hoon, and Sung Bum Pan, "Improvement of Fingerprint Verification by Using the Similarity Distribution", *Information Technology Convergence*. Springer Netherlands, pp. 203-211, 2013.
- [7] Dong, Jing et Tan, Tieniu. Security enhancement of biometrics, cryptography and data hiding by their combinations. In : *Visual Information Engineering*, 2008. VIE 2008. 5th International Conference on. IET, pp. 239-244. 2008.
- [8] Desai, Miss Jyotsna T., and Miss Shashikala R. Jaiswar. "Secure ATM System with Biometric Fingerprints Technology." *International Journal of Current Trends in Engineering and Research 2.6* (2016): 489 – 494.
- [9] Chergui Othaila, Meraoumia Abdallah, Bendjenna Hakim. "Enhancing security and privacy in e-banking transactions using palmprint features. A case study of biometric incorporated ATM", unpublished.
- [10] Chergui Othaila, Meraoumia Abdallah, Bendjenna Hakim. "Combining Biometrics With Watermarking To Enhance Copyright Verification", unpublished.

Conception et développement de mécanismes à base des systèmes multi-agents tolérants aux fautes pour l'Adaptation dynamique des réseaux virtuels

Noureddine HATTAB

Directeurs de thèse : Zahia GUESSOUM

Équipe :/

Résumé— Dans des environnements virtuels fortement dynamiques (Cloud), la tolérance aux fautes "SFT" doit être réalisée automatiquement et à la demande. En cas de défaillance ou de panne, des systèmes multi-agents tolérants aux fautes sont indispensables pour adapter ces mécanismes (schéma) aux changements inopinés de ces environnements non traditionnels, d'une manière distribuée et collaborative. Il s'agit de concevoir les mécanismes d'auto-organisation et de coopération entre les agents intelligents responsables de l'allocation et adaptation dynamique des ressources virtuelles (VMs).

I. INTRODUCTION

Quand on parle de SFT préventive (réplication et reprise) dans des environnements virtualisés (Cloud), on parle souvent d'un ensemble de concepts substantiels, profusément détaillés dans les environnements distribués, et qui représente des pierres angulaires pour toute prospection sur la tolérance dans les environnements cloud. Nous identifions et nous adoptons ici un ensemble de concepts retrouvés dans [9,10,11,13]: La diversité, l'adjudicateur (décideur) et ses modes de contrôle, la capacité de préservation de l'historique, le type et la corrélation à considérer de faute, erreur, défaillance, ainsi que la philosophie de recouvrement adoptée (point de reprise ou réplication), et finalement le mode de communication entre la requête et les nœuds virtuels.

Toutes ces concepts ont été l'objet d'une investigation approfondie, dans un but ultime, familiariser les différents modèles "SFT" de notre état de l'art.

II. PROBLÉMATIQUE ET ISSUES

Dans les différents systèmes distribués dynamiques et par extension virtualisés, la maîtrise de la SFT préventive passe inévitablement [9-10] par la maîtrise de l'action élémentaire de ses adjudicateurs, et la gouvernance de ses modes de contrôle adoptés (dynamiquement adaptatifs fortement recommandés)[9]. Dans la littérature ces décideurs adaptatifs adressent plusieurs questions, complexes des fois même paradoxales.

1) Connaissance continue des ressources disponibles de l'environnement virtualisé global, que ce soit inter-VM (nombre de VM) ou local intra-VM (CPU, Hdd, Power, etc), ainsi que les exigences d'utilisateurs (Requête) notariées par un contrat de garanti de service SLA.

2) L'état du schéma de recouvrement préconisé, en terme de choix de la Forme (Point de restauration ou réplication (active ou passive)) et en terme de quantification du contenu (nombre de répliques, l'intervalle du temps pour le point de restauration ...etc).

3) La capacité de mesurer la qualité (fiabilité) des modules de réplique pour un schéma de réplication donné (RA), dans un

intervalle de temps statique (système synchrone) ou dynamique (système asynchrone?). Ici le défi réclame que le temps nécessaire min-max pour un adjudicateur ne doit pas être fixe pour que le protocole de détection de faute déclare qu'une réplique est en panne ou non; mais soit dynamique en profitant de l'historique décisionnel précédent de l'ensemble de réplique [Pierre S 2016]

L'enjeu majeur attendu par l'adaptation (On-line) du SFT, dans les environnements virtualisés 'Cloud', est de regrouper toutes les défis et les mesures et de le prendre à la fois en considération (même s'ils ont contradictoire par nature) toute en estimons l'état actuel et continu de l'environnement; pour un deux raisons, d'un côté, maximiser respectivement le bénéfice/fiabilité du fournisseur et du consommateur de cloud, et de l'autre côté, minimiser respectivement les ressources occupées et non utilisables pour toute schémas SFT donné; et les pénalités qui résultent la transgression du 'SLA' d'un client certifié.

III. ETAT DE L'ART

Dans la littérature, plusieurs modèles et techniques ont été proposées afin d'affronter la défaillance et par conséquent rupture de service dans ces environnements distribués.

Cette section récapitule huit modèles (architectures) exclusivement destinés à la renforcement de la fiabilité et de la tolérance aux fautes préventives dans les environnements virtualisés.

Ces modèles sont alignés dans le tableau (Tab.1) par rapport aux concepts identifiés et expliqués antérieurement dans l'introduction.

IV. PERSPECTIVES

Dans les environnements Cloud, il n'existe pas une solution SFT génériques (hypothèse; objectif et métrique différents), qui s'adapte à toutes les situations et les scénarios.

De ce fait le travail convoité prochainement, est de proposer **une architecture Astucieuse SMA-SFT** en compagnie de stratégie de répliques adaptatives, qui adopte un mode de communication de type multi-casté i.e la relation de la requête avec le cloud et de type (1.*). Intérieurement la complexité est majeure, notamment lorsque on aborde la question de la consistance des répliques différents pour des requêtes différents avec un emplacement (VM) commun, en tenant en considération la QoS attendu par les différents clients notariés par ses propres SLAs.

On doit s'inspirer de l'ensemble des contributions de la communauté multiagent tolérante aux fautes; quant à affronter la faute dans ces environnement fortement dynamique à large échelle, on focalisons toute nos attentions sur les protocole de coopération et de coordination entre les agents intelligents . Les agents ici sont incarné soit au niveau des adjudicateurs, soit au niveau de répliques (primaire) ...etc. dans une vision global , chaque composant est un agent.

Critères	Attributs	Modeles /Architectures							
		1	2	3	4	5	6	7	8
formes de Diversité	conception	*	*	*	*	*	*	*	*
	Data			*	*				*
	Autres								
Adjudicateur Décideur	AT								
	Voter/Comp								
	Hypride	*	*	*	*	*	*	*	*
Mode de contrôle (Exécution)	Dynam	*	*	*	*	*	*	*	*
	Parallel								
Le niveau Système Considérer	Materiel				*				
	Logiciel (VM, VMM)	*	*	*	*	*	*	*	*
Type Fautes logiciel	Solide			*	*				*
	Softte	*	*	*	*	*	*	*	*
Architecture systems	Centralisée								
	Decentralisée								
Historique système (CCP)	Non								
	Oui (CS / Non)	*	*	*	*	*	*	*	*
Défaillance Type	Detectée	*		*	*				*
	Non		*			*	*	*	
Corrélation des fautes	Liée			*	*		*		*
	Non	*	*			*		*	
Errors Variants	IAW		*			*		*	*
	MCR	*		*	*		*		
Recouvrement	Reprise				*	*	*		*
	Réplication	*	*	*	*			*	*
Mode de communication	(1 pour 1)	*	*	*	*	*	*	*	*
	(1 pour n)								
	(n pour n)								

Tab.1 : Etat de l'art

V. CONCLUSION

Actuellement il existe de nombreux modèles SFT, avec une multitude de mécanismes, schémas pour augmenter a la fois la performance et la fiabilité. La plupart de ces environnements virtualisé actuel (cloud) adopte les méthodes préventive (La réplication, reprise).

Il reste encore de nombreux défis et challenges; surtout dans le volet d'adaptation. Dans le cadre de protocole de la détection de fautes parmi les répliques ou lors de l'élection de réplique primaire en case de panne de cette dernière. i.e l'interaction et la négociation entre l'adjudicateur et les répliques pose un problème de comparaison et de consistance (CCP). de ce fait ces protocoles doivent être améliorer et optimiser sa qualité de négociation par la coordination entre l'ensemble de répliques (agents) constituant un fragment; la notion d'agent ici stimule la capacité de prendre des action et des initiations sans retournée aux adjudicateurs, a la base de la situation courante de l'environnement, et les exigence approuvé par le SLA.

Peu de travaux qui prend en compte des SMA-Tolérant aux fautes comme un cadre de modélisation pour améliorer les protocole de coordination entre un fragment de répliques particulièrement lorsque le mode de communication adopté est de type multicast(1 ..n).

Références

- [1] Wenbing Zhao, P. M. Melliar-Smith, Louise E. Moser: **Low Latency Fault Tolerance System**. Comput. J. 56(6): 716-740 (2013)
- [2] Malik ,Huet F.Adaptive fault tolerance in real time cloud computing, Services (SERVICES), IEEE World Congress pp.280–287 :4–9 July 2011.
- [3] Ravi Jhawar and all; Fault Tolerance Management in Cloud Computing: A System-Level Perspective IEEE SYSTEMS JOURNAL, VOL. 7, NO. 2, JUNE 2013
- [4] Zhang Y, and all; .BFT Cloud: a byzantine fault tolerance framework for voluntary-resource cloud computing in cloud computing (CLOUD). In: Proceedings of the IEEE international conference on 4–9
- [5] Pranesh Das, Dr.Pabitra Mohan Khilar , VFT: **A Virtualization and Fault Tolerance Approach for Cloud Computing** Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).
- [6] S. Shen, A. Iosup, A. Israel, W. Cime, D. Raz, and D. Epema, An Availability-on-Demand Mechanism for Datacenters, IEEE/ACM Int. Symp. Clust. Cloud Grid Comput., pp. 495–504, 2015.
- [7] MOHAMMED AMOON – “Adaptive Framework for Reliable Cloud Computing Environment” *Digital Object Identifier/ACCESS.2016.2623633*. VOLUME 4, 2016.
- [8] Bashir Mohammed and all “An Integrated Virtualized Strategy for Fault Tolerance in Cloud Computing Environment”. 2016 Intl IEEE
- [9] L. Pullum, "software Fault-tolerant techniques and implementation" books Artech House computing library, 2004.ISBN 1-58053-137-7.
- [10] Laprie, J et al , "Architectural Issues in software Fault Tolerance", Software Fault Tolerance, New York: John Wiley & Son 1995 pp 47-80
- [11] A. Avizienis, et all “Basic concepts and taxonomy of dependable and secure computing,” Dependable and Secure Computing, IEEE Transactions on, vol 1, no. 1, 2004.
- [12] G. Latif-Shabgahi, et all. A taxonomy for software voting algorithms used in safety critical systems. *IEEE Transactions on Reliability*, 53(3):319–328, September 2004.
- [13] J. Liet all "Fault tolerance and scaling in e-Science cloud applications: observations from the continuing development of MODISAzure." In e-Science (e-Science), 2010 IEEE 6th International Conference on, pp. 246- 253. IEEE, 2010
- [14] Eric Bauer and Randeep *Reliability and Availability of Cloud Computing*, First Edition. Adams. © 2012 Institute of Electrical and Electronics Engineers. Published 2012 by John Wiley & Sons, Inc.

Apport des informations imparfaites dans les systèmes de détection d'intrusions

Abdelkader ALEM

Youcef DAHMANI, Allet HADJALI

Résumé— Dans notre contexte, nous nous intéressons à la modélisation des systèmes à paramètres imparfaits (informations imparfaites) qui peut être efficace pour détecter les problèmes (intrusions réseaux) réels. Par nature, les informations des systèmes complexes sont imparfaites (c-a-d imprécises, incertaines, etc). Ainsi, pour expliquer ou anticiper les états de ces systèmes, un formalisme de représentation et de manipulation des informations paraît nécessaire, parmi les outils pour considérer les imperfections et prendre des décisions. : La théorie de l'incertain (les fonctions de croyance et les réseaux bayésiens), la théorie de l'imprécise (la logique floue). L'objectif est de proposer des solutions aux problèmes complexes et de développer une architecture de système de détection d'intrusion.

I. INTRODUCTION

La sécurité informatique est devenue une obligation pour toute organisation pour faire face aux attaques afin de minimiser les risques. La sécurité informatique est la protection de l'information et des systèmes d'information contre les accès, la perturbation afin de garantir la confidentialité, l'intégrité et la disponibilité.

Les systèmes et réseau informatique contiennent diverses formes de vulnérabilité. Pour faire face à ces problèmes de sécurité informatique, différents mécanismes ont été mise en place pour prévenir toute sorte d'attaque comme les pare-feu, l'authentification...etc. malheureusement ces mécanismes ont des limites ou certain types d'attaques peuvent les contourner pour nuire la politique de sécurité.

Pour faire face à ce problème, un nouveau concept s'appelle système de détection d'intrusion a été introduit comme une seconde ligne de défense pour protéger contre les activités malicieuses. La détection d'intrusions a pour objectif de détecter toute violation de la politique de sécurité en vigueur sur un système d'information. Il existe deux principales techniques pour la détection de l'intrusion : la détection des anomalies (approche comportementale) et la détection des malveillances (approche par signatures). La détection des anomalies consiste à construire un modèle identifiant les comportements déviants du modèle comportemental normal. La détection des malveillances, quant à elle, compare le comportement d'utilisation du système avec des signatures d'attaques connues.

Les systèmes de détection d'intrusions par signature fondent leur détection sur la reconnaissance, dans le flux d'évènements générés par une ou plusieurs sondes de signatures d'attaques qui sont contenues dans une base de signatures. Les deux principaux avantages de solutions suivant l'approche par signature sont la pertinence des alertes et une certaine facilité de mise en place de l'IDS. Cependant, cette approche présente certaines limites, dont la plus gênante, est de ne pouvoir détecter que les attaques connues.

L'approche comportementale est basée sur l'observation des sujets en situation normale. Des statistiques sont établies sur un certain nombre d'informations les concernant, et elles sont apprises en observant le système pendant une période donnée. Les détecteurs utilisant cette approche sont par la suite capables de repérer des comportements anormaux. Une alerte est levée dès lors qu'une déviation est constatée entre le comportement statistique appris et le comportement observé. Cette méthode est intéressante parce qu'elle semble permettre de détecter des nouvelles attaques. En effet, il est fort probable qu'une nouvelle attaque dévie du comportement habituel des utilisateurs. Cependant, elle souffre également de deux gros inconvénients. Il n'y a pas réellement de diagnostic établi. On se contente de remarquer que le comportement est anormal. Il faut donc faire une analyse complémentaire pour diagnostiquer l'attaque. De plus le nombre de faux positifs, de fausses alertes, est assez élevée car il n'y a pas que les attaques qui entraînent des comportements inhabituels. Ainsi, l'administrateur qui a pour tâche d'analyser et de prendre des décisions nécessaires, se retrouve rapidement débordé, et laisse certainement passer des alertes suspectes sans les analyser.

II. PROBLEMATIQUE

Les systèmes de détection d'intrusions (IDS) sont généralement considérés comme une seconde ligne de défense pour protéger contre les activités malicieuses.

La détection des tentatives d'attaques sur un réseau est une problématique très importante dans le domaine de la sécurité informatique. Les technologies classiques de protection des réseaux de type Firewall filtrant sont en effet inefficaces contre la plupart des attaques actuelles. Aussi sont apparus de nouveaux équipements réseaux pour prendre en compte ces carences, les NIDS, systèmes de détection d'intrusions réseaux, dont le but est de détecter les tentatives d'attaque qu'un firewall ne peut pas bloquer. Malheureusement, en pratique, les NIDS génèrent tellement d'alertes sur un réseau important qu'il en devient très difficile de déterminer celles générés par une attaque réelle. La corrélation d'alertes est un mécanisme indispensable pour la réduction du volume important des alertes et pour la détection des attaques coordonnées et complexes. Les approches existantes soit se basent sur des connaissances d'experts, soit utilisent des mesures de similarité simples qui ne permettent pas de détecter des attaques complexes. Elles souffrent également d'une complexité de calcul très élevée due, par exemple, à un grand nombre de scénarios possibles pour détecter une attaque coordonnée.

Dans notre contexte, nous nous intéressons à l'utilisation des informations imparfaites qui peut être efficace pour détecter les problèmes réels. Par nature, les informations des systèmes complexes sont imparfaites - c.-à-d., imprécises,

incertaines, etc. Ainsi, pour expliquer ou anticiper les états de ces systèmes, un formalisme de représentation et de manipulation des informations paraît nécessaire. Si les simulations numériques permettent de modéliser un système complexe (la fusion d'informations qui consiste à combiner des informations hétérogènes et tenir compte des spécificités des données issues de plusieurs sources afin d'améliorer la prise de décision), elles doivent être accompagnées d'un outil pour considérer les imperfections et prendre des décisions. Parmi ces outils : (i) la théorie de l'incertain (les fonctions de croyance et les réseaux bayésiens), (ii) la théorie de l'imprécise (la logique floue) qui sont particulièrement bien adaptés aux systèmes complexes et à la modélisation des différents types d'imperfection.

L'objectif est de proposer des solutions aux problèmes complexes et de développer un système de détection d'intrusion réseaux efficace en minimisant le taux des fausses alertes.

III. TRAVAUX REALISES

Notre première proposition a été mise en œuvre en utilisant la théorie de Dempster-Shafer afin de combiner les résultats de plusieurs classificateurs.

Dans la deuxième proposition on a construit un modèle multi-niveau. L'idée est d'utiliser deux classificateurs différents d'une manière itérative, ou chaque itération représente un niveau dans le modèle créé. Les sortie de chaque classificateur sont fusionnés en utilisant la règle de combinaison conjonctive, nous ajoutons un nouveau niveau chaque fois que le conflit entre classificateurs atteint le seuil. Pour construire notre modèle, nous avons utilisé "Fuzzy Unordered Rule Induction Algorithm" et "Random Forests" comme classificateurs, et nous avons utilisés la logique floue pour fuzzifier les sorties des deux classificateurs et initialiser les masses de l'étape de fusion qui est basée sur la théorie de l'évidence de Dempster Shafer (TDS).

A. Travaux SITIS 2016

Un papier intitulé : *On the use of Belief Functions to improve High Performance Intrusion Detection System* a été présenté dans la conférence intitulé : *International Conference on Signal Image Technology & Internet Systems (STIS)*. La principale contribution porte sur l'utilisation de la règle orthogonale de Dempster pour combiner les résultats de deux sources de données dans les IDSs .

Dans ce papier [1], un IDS hybride a été proposé basé sur la théorie des fonctions de croyance, le system proposé contient trois niveaux (Fig.1). On trouve deux différents classificateurs (SVM et les Réseaux Bayésiens) qui forme le premier niveau, le deuxième niveau représente la phase de la fuzzification des sorties de chaque classificateurs en se basant sur la logique floue. En fin, un module de fusion basé sur la règle de combinaison de Dempster pour combiner les résultats issus du deuxième niveau par la théorie des fonctions de croyance (règle de combinaison conjonctive de Dempster).

Dans ce travail, notre approche a été testée en utilisant un corpus des attaques réels, la base d'apprentissage contient 98995 connexions avec 24 attributs et la base de test contient 69015 entées, La TABLE I présente la distribution des connexions dans l'ensemble de données de test et d'apprentissage.

TABLE I DISTRIBUTION OF CONNECTION CATEGORIES IN THE DATASET

Classe	Données d'apprentissage		Données de Test	
	Nombre	%	Nombre	%
Normal	55344	55,90	61378	88,93
Anormal (Attaque)	43651	44,10	7637	11,07
Total	98995	100	69015	100

La performance d'un IDS est mesurée par sa capacité de classer chaque connexion dans la bonne catégorie. Les indicateurs de performance les plus utilisés pour évaluer les

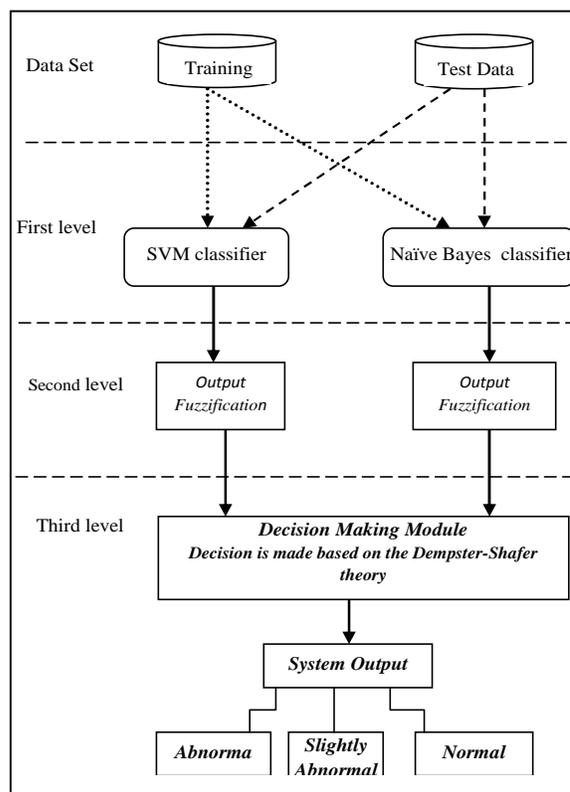


Figure 1 Structure générale de notre Modèle.

systèmes de détection d'intrusion sont :

- Exactitude = $(D+A)/Total$;
- Faux positif = $B/(B+A)$;
- Faux négatif = $C/(C+D)$;

Ou :

- A : Les connexions anormales détectés comme anormales.
- B : Les connexions normales détectées comme anormales.
- C : Les connexions anormales détectées comme normales.
- D : Les connexions normales détectées comme normales.

Pour tester notre solution, l’outil WEKA DataMining est utilisé. La Figure 2 et la Table II illustrent les résultats d’expérimentation.

TABLE II RESULTATS D’EXPERIMENTATION

	SVM	Naïf Bayes	Proposed Approach (DSIDS)
Accuracy	99,12%	97,81%	99,73%
Detection rate	99,71%	99,71%	99,71%
False alarm rate	7,13%	16,33%	1,47%

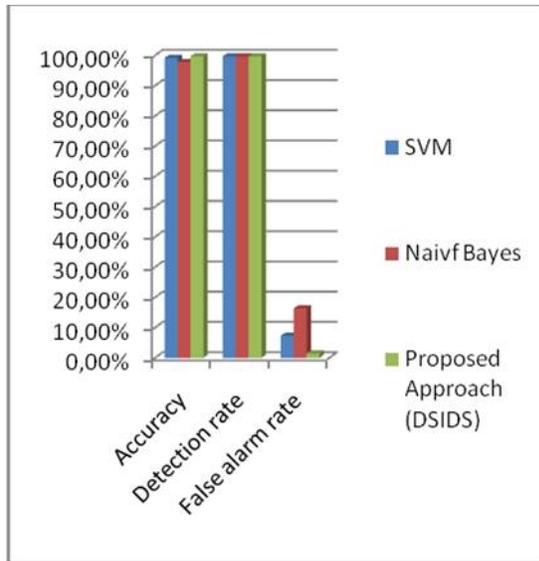


Figure.2 Comparaison entre notre approche et autres classifieurs.

B. Travaux IJSN 2017

Un papier intitulé : *A new High Performance Intrusion Detection System* à été soumis pour publication à un journal international intitulé : *International Journal of Security and Network*. La principale contribution porte sur l’utilisation des fonctions de croyances pour construire un modèle hiérarchique et hybride.

L’approche proposée dans [2] est de développer un IDS hybride, multi-niveaux. L’expérimentation avec l’ensemble de données KDD’99 a montré les performances de notre modèle à travers trois critères importants qui sont: un taux de détection élevé et un taux de faux positifs faible. En utilisant même ensemble de données et même classifieurs, notre modèle à présenté des meilleures performances par rapport au travail présenté dans [3].

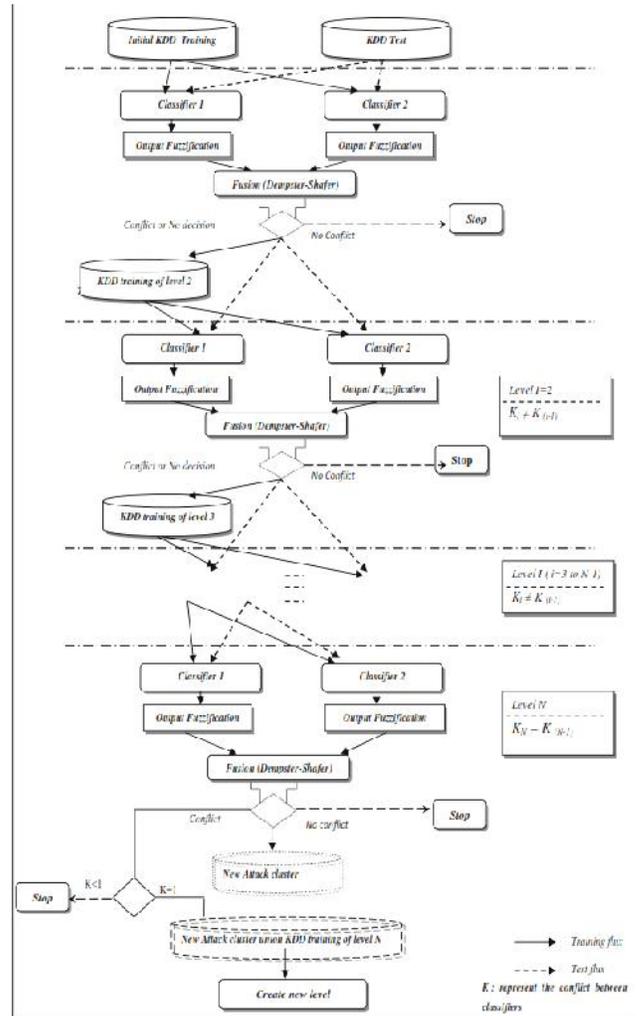


Figure.3 Structure générale de notre deuxième Modèle.

IV. PERSPECTIVE

Ce que nous envisageons de faire durant l’année 2016/2017 : la finalisation de la rédaction de la thèse, L’attente des commentaires des sur le deuxième papier.

V. CONCLUSION

Notre contribution est basée sur l’utilisation des fonctions de croyance et la règle de Dempster pour considérer les imperfections et prendre des décisions. Cette proposition devrait améliorer les performances des IDS en minimisant le taux des faux positifs et maximiser l’exactitude.

Références

[1] A. ALEM, Y. DAHMANI, and A. HADJALI.: On the use of Belief Functions to improve High Performance Intrusion Detection System. International Conference on Signal Image Technology & Internet Systems (STIS). 2016.

[2] A. ALEM, Y. DAHMANI, and A. HADJALI.: A new High Performance Intrusion Detection System. International Journal of Security and Network. 2017 “unpublished”.

[3] A. AHMIM, N. GHOUALMI-Zine, A new adaptive intrusion detection system based on the intersection of two different classifiers, international journal of security and network, vol.9, No. 3, pp 125-132. 2014

[4] <http://www.cs.waikato.ac.nz/ml/weka/> accessed 21/09/2016.

Exploration et exploitation des aspects émotionnels chez l'apprenant en vue d'une définition d'un profil émotionnel pour un meilleur apprentissage en e-Learning

Tarek Boutefara

Latifa Mahdaoui

Equipe : KMEELEA, Laboratoire : RIIMA

Adresse : BP N°32, El-Alia, 16111, Bab-Ezzouar, Alger.

Résumé—Pendant la sixième année de notre thèse, nous avons progressé en respectant l'échéancier fixé. Les progressions ont été réalisées sur les deux plans : technique et théorique. Sur le plan technique, une première version expérimentale de la plateforme flexible, modulaire et extensible qui devrait supporter l'intégration des différentes propositions effectuées durant la thèse est disponible. Sur le plan théorique, nous avons continué sur notre proposition majeure : construire un processus sur trois étapes (avant, durant et après l'activité pédagogique) pour la prise en compte des émotions sur les plateformes e-Learning. Nous avons pu travailler sur une révision de l'approche FEASP pour pouvoir l'intégrer au niveau de création et de déroulement des ressources pédagogiques; un composant qui représente la dernière brique dans la construction du processus visé.

I. INTRODUCTION

Notre thèse est intitulée " Exploration et exploitation des aspects émotionnels chez l'apprenant en vue d'une définition d'un profil émotionnel pour un meilleur apprentissage en e-Learning ". Après avoir réalisé notre étude théorique, nous avons trouvé que c'était nécessaire de réagir sur deux niveaux principaux : le niveau théorique et le niveau technique.

Le niveau théorique est le niveau abstrait où on discute l'exploration et la prise en compte des émotions des apprenants sur les plateformes e-Learning. Ces traitements sont définis sur trois phases différentes.

- **Avant l'activité pédagogique** : par le concepteur pédagogique.
- **Pendant l'activité pédagogique** : par la détection de l'état émotionnel de l'apprenant pendant l'activité pédagogique.
- **Après l'activité pédagogique** : pour pouvoir améliorer les supports pédagogiques.

Sur le niveau technique, deux solutions sont présentes. La première solution consiste à développer des outils dédiés à la détection de l'état émotionnel de l'apprenant. L'inconvénient majeur est le fait que ces systèmes sont, généralement, fermés, ainsi, des enjeux majeurs du domaine e-Learning telles que la réutilisation et l'interopérabilités des ressources seront perdus.

La deuxième solution consiste à exploiter les outils existants dans la conception des supports pédagogiques émotionnels. L'idée est de garder la réutilisabilité et l'interopérabilité des ressources avec les systèmes existants.

L'inconvénient de cette solution réside dans la compréhension de la dimension émotionnelle de ces outils.

Durant la sixième année, nous avons réalisé des progrès en suivant l'échéancier fixé. Ainsi, dans cet rapport, nous allons rappeler la problématique fixée pour cette année ensuite nous allons présenter les travaux réalisés durant l'année. Nous terminerons par exposer nos perspectives pour d'autres travaux de recherche.

II. PROBLÉMATIQUE

La personnalisation sur les plateformes e-Learning est l'un des axes de recherche les plus actifs. Elle vise à fournir un contenu, un enchaînement ou une configuration adaptés aux besoins de l'apprenant.

Pour garantir une adaptation plus approfondie, il faut voir plus loin que les styles d'apprentissage et les préférences utilisateur. En effet, la prise en compte de l'état émotionnel de l'apprenant est un facteur clé pour une meilleure adaptation.

Notre travail s'inscrit dans cette vision. Ainsi, nous travaillons sur deux actions principales : explorer la dimension émotionnelle des outils et des techniques utilisés dans le cadre de l'e-Learning, et exploiter cette dimension pour offrir une meilleure expérience d'apprentissage.

Le développement de ces deux actions nous a conduit à travailler sur les deux niveaux d'abstraction et sur les trois phases indiqués dans l'introduction de ce rapport.

Durant les trois dernières années (2013/2014, 2014/2015 et 2015/2016), nous avons réalisé un ensemble de travaux et des expérimentations qui s'inscrivent dans cette vision. Les travaux réalisés, brièvement, sont :

- Un outil pour récupérer l'état émotionnel de l'émotion sous la forme d'un feedback en utilisant les émojis.
- L'extension de la norme IMS-LD pour supporter la dimension émotionnelle.
- Des expérimentations sur l'extraction de l'état émotionnel à partir du texte. Nous avons pu l'appliquer sur les traces textuelles laissées par les étudiants sur la plateforme Moodle.

En travaillant sur ces projets, nous avons rencontrés plusieurs difficultés. Ces observations ont constitué le fondement de nos travaux durant cette dernière année.

III. TRAVAUX RÉALISÉS

A. Exploration de la dimension émotionnel de quelques outils

Nous avons continué nos efforts d'exploration pour mieux comprendre la dimension émotionnelle des outils pédagogiques et collaboratifs présents sur les plateformes e-Learning. Nos travaux se sont concentré sur deux outils et se sont résumé en deux projets. Le premier projet visait à étudier l'effet des environnements e-Learning 3D et la présence virtuelle qu'ils offrent sur l'état émotionnel d'isolement. Ce dernier est considéré comme un inconvénient de l'utilisation des supports textuels en e-Learning.

Le projet a permis de réaliser une classe virtuelle pour l'enseignement du module algorithmique pour les étudiants en première année universitaire. Cette classe a été, ensuite, testé par 20 étudiants. Le test a pris la forme d'une expérimentation suivi d'un sondage pour récupérer les feedbacks nécessaires de la part des étudiants. Parmi les résultats obtenus, nous citons que 90 % des étudiants ont rouver la présence de l'enseignant très encourageant pour suivre le cours dans l'environnement 3D. Ce travail a été accepté comme « Short Paper » dans le cadre de la conférence CELDA2017.

Notre deuxième projet consistait à étudier les possibilités offertes par les annotations laissées par les apprenants comme source d'extraction des indices émotionnels. Nous voyons ce projet comme une suite à l'outil de feedback émotionnel basé sur les émoticônes. Néanmoins, les possibilités offertes par les annotations (qui sont localisées et dotées d'un aspect graphique plus riche) sont plus promoteuses mais plus difficiles à prouver et à modéliser. Ce travail a donné comme fruit un article qui a été soumis à la conférence ACIT2017.

B. Construction d'une plateforme e-Learning capable de supporter l'aspect émotionnel

Nous avons travaillé sur deux prototypes qui visent à résoudre les problèmes de support de l'aspect émotionnel.

La première solution consiste à opter pour une architecture basée sur le smicroservices. Un microservice est un composant logiciel autonome qui assure une seule fonctionnalité. Cela implique une modularité maximale et un couplage très faible. Ces caractéristiques sont facilement applicables dans les environnements e-Learning. En effet, ces derniers sont, généralement, conçus autour de deux notions indépendantes : les ressources pédagogiques et les apprenants. Le prototype expérimental est prêt et un article en cours de rédaction.

La deuxième approche vise à résoudre le problème d'hétérogénéité constaté avec les différences notées entre les différentes sources des indicateurs émotionnels. La solution que nous proposons est de concevoir un système sensible à l'état émotionnel en se basant sur un système multiagents. Chaque agent se charge de la collecte d'une famille homogène des indicateurs pour ensuite faire la projection sur un modèle commun entre tous les agents. Ce travail est en cours de finalisation avec un autre doctorant (M. Attia Mourad, école doctorale STIC, ESI) qui se spécialise dans le domaine des systèmes multiagents.

C. De l'extension IMS-LD vers une framework de construction des supports pédagogiques émotionnels

L'extension proposé précédemment permet de définir explicitement une dimension émotionnelle pour le support pédagogique développé. Cette dimension est répartie sur les différents éléments du support pédagogique par l'introduction de nouvelles notions comme : objectif émotionnel, modèle émotionnel, condition émotionnel et opérateur émotionnel. Cette extension a donné lieu à un outil d'exécution (player) qui peut l'interpréter.

L'étape suivante consistait à travailler sur un système auteur qui supporte cette extension (même en arrière plan). Ains, il était nécessaire de proposer un modèle émotionnel prédéfini qui encadre les enseignants durant la construction de ses supports pédagogiques émotionnels. Pour cela, nous avons choisi le modèle FEASP qui est l'un des rares modèles proposés dans ce cadre. Le prototype est en cours de développement en utilisant l'API P5 (la version JavaScript du langage Processing). Une fois accomplie et évaluée, nous aurons complété notre processus cible en offrant une étude sur les trois phases à savoir : la prise en compte de l'état émotionnel de l'apprenant avant, durant et après l'activité pédagogique.

IV. PERSPECTIVES

Tout au long de nos travaux de recherche, une question ne cessait de se répéter et à laquelle tous les travaux et les expérimentations réalisés nous conduisaient : "Qu'est ce que l'émotion ? Et quelle est la meilleure représentation informatique (technique) de l'émotion ?

Nous posons cette question en accord avec la confirmation de Pr. Rolf Pfeifer qui dit : « If we want to understand the function of the brain (or the control in the case of robots), we must understand how the brain is embedded into the physical system, and how the organism interacts with the real world. »

Dans le cadre de notre travail, cette affirmation sera adaptée pour cibler la compréhension du fonctionnement des émotions. Ainsi, nous avons entamer un travail pour construire un réseau de neurone émotionnel. A l'heure actuelle, un et un seul travail a été réalisé dans ce sens. Nous visons à réaliser une implémentation fidèles des aspects physiques liés au phénomène des émotions et leurs influences dans le cerveau humain. Cet objectif est fixé comme une perspective à long terme.

V. CONCLUSION

Dans ce rapport annuel, nous avons présenté notre état d'avancement par exposer les différents travaux réalisés durant l'année universitaire 2016/2017. Pour faire, nous avons donné introduit les travaux réalisés durant les dernières années ainsi que notre problématique.

Dans ce rapport, nous nous sommes focalisés sur les grandes lignes des travaux réalisés et en cours de réalisation et particulièrement la construction d'une framework pour la conception des supports pédagogiques. En association avec les projets réalisés dans ce sens, nous pouvons voir l'accomplissement du processus que nous avons défendu durant les dernière années et nous aurons ainsi accompli le noyau fondamentale de notre contribution.

Portail Sémantique pour la Gestion de Connaissances Destiné au Corpus (en langue arabe) des Hadiths du Prophète.

Meftah DAHMOUNI

Directeur de thèse : Kamel BOUKHALFA

Co-directeur : Hassina ALIANE

Résumé— suite aux orientations de la commission de suivi, qui a valorisé les travaux que nous avons accomplis (l'ontologie, développée sur la base d'une nouvelle classification proposée, l'application réalisée pour la personnalisation et l'exploitation de cette ontologie, ainsi que l'approche proposée pour extraire les sujets d'El Hadith, nous avons élaboré un article relatif à la classification proposée, et également nous avons commencé à programmer l'approche d'extraction des sujets que nous avons proposée.

I. INTRODUCTION

Il existe plusieurs types de classifications des Ahadiths, mais chacune d'elles comporte des inconvénients majeurs, à l'instar du non couverture de la totalité du corpus des Ahadiths. Pour cela, nous avons effectué une étude critique des classifications existantes, afin d'extraire les points forts et les points faibles de ces classifications, ainsi nous avons montré notre classification proposée. Cette classification fera l'objet d'exploitation dans la recherche des sujets dans le corpus Ahadiths, en utilisant l'approche que nous avons proposée l'année précédente, avec des améliorations qui ont été ajoutés.

II. PROBLEMATIQUE

L'approche que nous avons proposée et en cours de réalisation, elle rentre dans le cadre de notre travail qui consiste à réaliser un portail du web sémantique [1], [2] relatif au corpus des Ahadiths du prophète. Ce portail est destiné aussi bien au grand public qu'aux spécialistes du domaine (savants musulman, chercheur, ...). Sachant qu'il comporte :

- Une ontologie [3] des concepts (sujets) généraux des Ahadiths, qui représente notre classification,
- Une interface de navigation sémantique.

Cette dernière permet la mise à jour de l'ontologie (l'ajout, la modification et la suppression des Ahadiths) et sa personnalisation (ajout, modification et suppression des classes) et la recherche des Ahadiths par un sujet donné. Cette dernière (personnalisation) permettant à l'utilisateur (public ou expert) pour l'améliorer.

Egalement, cette interface permettant l'extraction des sujets dans un Hadith donnée, via l'approche que nous avons proposée.

III. TRAVAUX REALISES

En se basant sur les critères d'évaluation, nous avons les points suivants à présenter :

- A. *Avancement différentiel: Niveau d'avancement par rapport à l'année précédente*
 - Nous avons élaboré un article en montrant les points forts de notre classification (notre ontologie) basé sur une étude comparative des classifications existantes que ce soient des classifications informatiques (ontologies [4] [5], ...) ou des classifications qui existent dans les livres des sciences islamiques (livres des Ahadiths [6] [7] ...).
 - Nous avons amélioré notre approche d'extraction des sujets des Ahadiths que nous avons proposé avant, en introduisant la levée d'ambiguïté pour les concepts qui peuvent prendre plusieurs significations. L'approche que nous avons proposée, permet préalablement de trouver le nombre optimal des sujets qu'il comporte El Hadith, en le divisant (i.e. El Hadith) en un ensemble de parties suivant des règles que nous avons définies. De plus, nous avons commencé à la programmer avec le langage python.
- B. *Publications et communications. Contributions avec production d'articles*

Nous avons participé aux deux conférences :

1. ASD 2016 Annaba (Papier cours).

Le titre de l'article : «Ontologie générique des concepts des Ahadiths, El Nabawia El Charifah ».

Dans ce papier, nous avons présenté les différents types de classifications présentées dans les livres des sciences islamiques et qui sont présentées également par des ontologies. Ensuite, nous avons donné une description générale de notre ontologie, et enfin nous avons présenté une première application (Mise à jour, personnalisation et recherche) pour exploiter cette ontologie.

2. *COSI 2016 Sétif (Poster).*

L'article intitulé : « Classification des Ahadiths à base des concepts d'une ontologie ».

Dans ce poster, nous avons donné préalablement les principes suivis pour construire notre ontologie. Par la suite, nous avons montré la possibilité d'utiliser cette ontologie comme module dans un Big Data des sciences islamiques, afin d'extraire les connaissances contenues dans les livres islamiques ou des encyclopédies (y compris les livres des Ahadiths). Ces connaissances sont présentées sous forme des bases de données, sous forme des fichiers XML, sous forme des fichiers textuels ou encore sous forme des sites web.

IV. PERSPECTIVES

Il nous reste à accomplir :

- Finalisation de la programmation de l'approche d'extraction des sujets en tenant compte des améliorations proposés l'année précédente,
- Evaluation de l'approche réalisée avec des experts en sciences islamiques (en particulier avec ceux du domaine des Ahadiths),
- Rédaction de la thèse.

V. CONCLUSION

La réalisation de notre approche, nous permettra de la mettre en évaluation auprès des experts du domaine des Ahadiths.

Références

- [1] T. Berners-Lee, J. Hendler, O. Lasilla, "The Semantic Web". In Scientific American, May 2001.
- [2] M. Şah, W. Hall, "Building and Managing Personalized Semantic Portals. In: Proceedings of International World Wide Web Conference", pp. 1227–1228, 2007.
- [3] M. Uschold, M. Gruninger, *Ontologies: Principles, Methods and Applications, Knowledge Engineering Review*; Vol. 11 N°2, 1996.
- [4] F. Harrag, A. Alothaim, A. Abanmy, F. Alomaigan, S. Alsalehi, "Ontology Extraction Approach for Prophetic Narration (Hadith) using Association Rules". *International Journal on Islamic Applications in Computer Science And Technology*, 2013.
- [5] M. G. Al-Masri, "An Ontology Based Approach to Enhance Information Retrieval from Al-Shamelah Digital Library". Thèse magistère, May 2015.
- [6] A. Ibn Hambel, "Musned El Imam Ahmed", établissement El Ressalah, 1996.
- [7] I. el Bokhari, *Sahih El Boukhari*. Traduit par Harakat Ahmed, El maktaba El Asriyah, Liban, Eème édition, 2003

PhD Plan

Name: Mohammed NAEM
Department:
Doctoral Programme: Computer Science
Title of PhD project: Internet of things Quality of Service Optimization Model

1. Description of the PhD thesis

a) Purpose

The purpose of the study is to provide a well suited Quality of Service Optimization Model taking into account the architecture, parameters, characteristics service specifications and metrics needed by certain internet of things application. It aims to presents problems and actual solutions in the literature about the quality of service in IoT, in an attempt to ensure necessary levels of service and performance of IoT application services.

b) Problem description

Because of the increasing use of IoT applications there are very few parameters that exist to check and ensure the quality of IoT implemented devices and machines, QoS schemes and parameters used in the traditional networks such as bandwidth, latency, jitter and loss of data is not sufficient to achieve the QoS in IoT environment it can be applicable to part of IoT network based on the underlying access network, however QoS in IoT parameters are viewed from various perspectives and dimensions such as the enabling technologies, users and application domain to provide the required, agreed and acceptable service quality for the users on specific IoT application domain.

c) Background and state of the art

IoT is the current wave of the Internet today, the goal of the internet of things is to enable things to be connected anytime, anyplace, with anything and anyone ideally using any path/network and any service, thereby QoS support for IoT is needed.

Some researches done already in QoS for M2M, However researchers should focus on IoT performance, optimization, quality (QoS, QoE), trust, safety, privacy and security.

Further presenting a generic conceptual model will be the contribution of this research as Internet of things Quality of Service Optimization Model

d) List of relevant literature

- 1-Chun-Wei Tsai • Chin-Feng Lai • Athanasios V. Vasilakos(2014). *Future Internet of Things: open issues and challenges*, Computer Networks.
- 2-Atzori, L., Iera, A., & Morabito, G. (2010). *The internet of things: A survey*. Computer Networks, 54(15), 2787–2805.
- 3-Katasonov, A., Kaykova, O., Khriyenko, O., Nikitin, S., & Terziyan, V. Y. (2008). *Smart semantic middleware for the internet of things*. In *Proceedings of the international conference on informatics in control, automation and robotics, intelligent control systems and optimization*, pp. 169–178..
- 4-de Saint-Exupery, A. (2009). *Internet of things strategic research roadmap*. In *Technical report, European research cluster on the internet of things*.
- 5-Lo´pez, T. S., Ranasinghe, D. C., Patkai, B., & McFarlane, D. C. (2011). *Taxonomy, technology and applications of smart objects*. Information Systems Frontiers, 13(2), 281–300.
- 6-Aggarwal, C. C. (Ed.). (2013). *Managing and mining sensor data*. Berlin: Springer.
- 7-Pan, J., Paul, S., & Jain, R. (2011). *A survey of the research of future internet architectures*. IEEE Communications Magazine, 49(7), 26–36.

8-Thiago Teixeira, Sara Hachem, Val'erie Issarny, and Nikolaos Georgantas .(2011).Service Oriented Middleware for the Internet of Things: A Perspective, 4th European Conference, Service Wave Poznan, Poland.

9-Andreas Metzger and Clarissa Cassales Marquezan.(2011). Future Internet Apps: The Next Wave of Adaptive Service-Oriented Systems?, 4th European Conference, Service Wave Poznan, Poland.

10-Johannes Riedl, Kolja Eger, Werner Mohr, and Ludwig Karg.(2011). ICT Enablers for Smart Energy, 4th European Conference, Service Wave Poznan, Poland.

2. Time Schedule

First Year 2016/2017			Timeline: 03/30/2017 - 12/04/2017
Task(s)			
Duration (days)	Start Date	End Date	Description
1	03/30/2017	03/31/2017	Arriving
5	04/02/2017	04/06/2017	First Year Registration
30	04/07/2017	05/18/2017	Looking for an interested research area
10	05/20/2017	06/02/2017	Determination of interested area(IoT)
20	06/04/2017	06/30/2017	intensive Reading to explore IoT open Issues
30	06/09/2017	07/20/2017	Discovering QoS in IoT related work
80	07/21/2017	11/09/2017	Working to identify problems and suggested contribution
90	08/01/2017	12/04/2017	Planning to prepare a survey paper about QoS in IoT
32	07/01/2017	08/14/2017	suggestion of thesis table of contents
Second Year 2017/2018			Timeline: 10/03/2017 - 11/27/2018
Task(s)			
Duration (days)	Start Date	End Date	Description
15	10/03/2017	10/23/2017	First year Progression Report
3	10/24/2017	10/26/2017	Second year registration
17	10/29/2017	11/21/2017	Finishing full research proposal
60	11/22/2017	02/13/2018	Completion of a survey paper (QoS in IoT)
45	02/14/2018	04/17/2018	Publication process and modifications
60	02/14/2018	05/08/2018	Formulation of chapters(1.2) of the thesis
80	05/01/2018	08/20/2018	Drafting phase of methodologies used and contribution
60	08/01/2018	10/23/2018	Preparation of implementing plan and tools
10	10/24/2018	11/06/2018	Evaluation of the current progression
25	10/24/2018	11/27/2018	Participation in a scientific conference

Third Year 2018/2019			Timeline: 10/01/2018 - 10/25/2019
Task(s)			
Duration (days)	Start Date	End Date	Description
15	10/01/2018	10/19/2018	Second year progression report
5	10/21/2018	10/26/2018	Third year registration
60	10/23/2018	01/14/2019	Experimentations and implementations results
60	01/05/2019	03/29/2019	Preparing and publishing of scientific contribution paper
30	03/15/2019	04/25/2019	Formulation of the third and fourth chapters of the thesis
15	04/25/2019	05/15/2019	Participation in a scientific conference
30	05/16/2019	06/26/2019	Evaluation of contribution and results
30	06/29/2019	08/09/2019	Formulation of the final chapters of the thesis
30	08/11/2019	09/20/2019	Reviewing the thesis
15	10/01/2019	10/21/2019	Third Year Progression Report
5	10/20/2019	10/25/2019	Last Registration

Diagrams of Schedule: First Year:

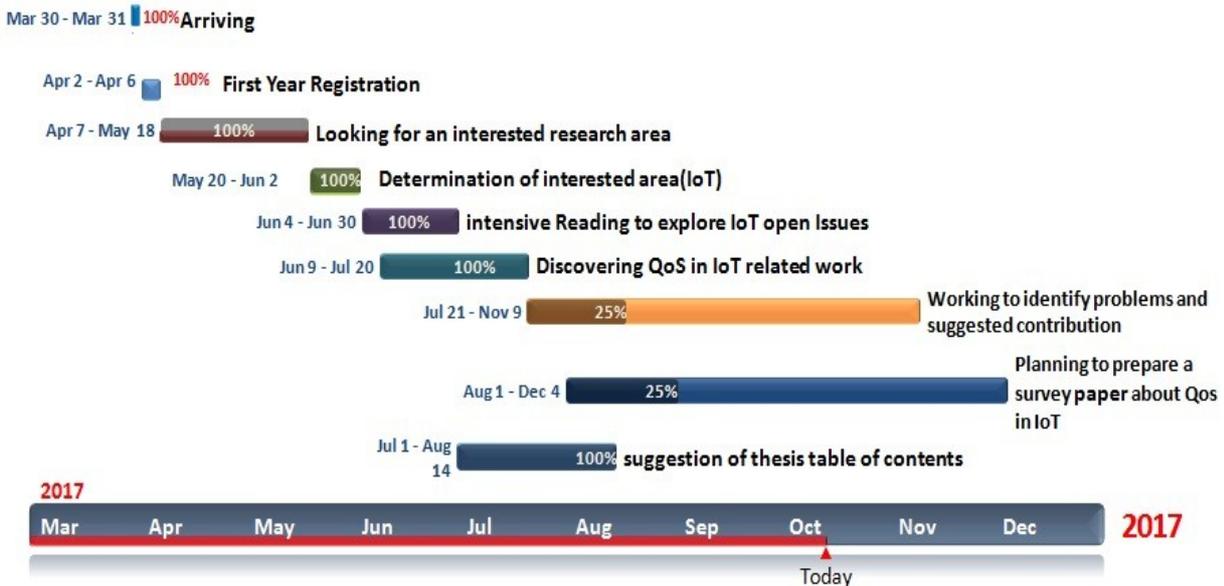


Fig:1

Second Year:

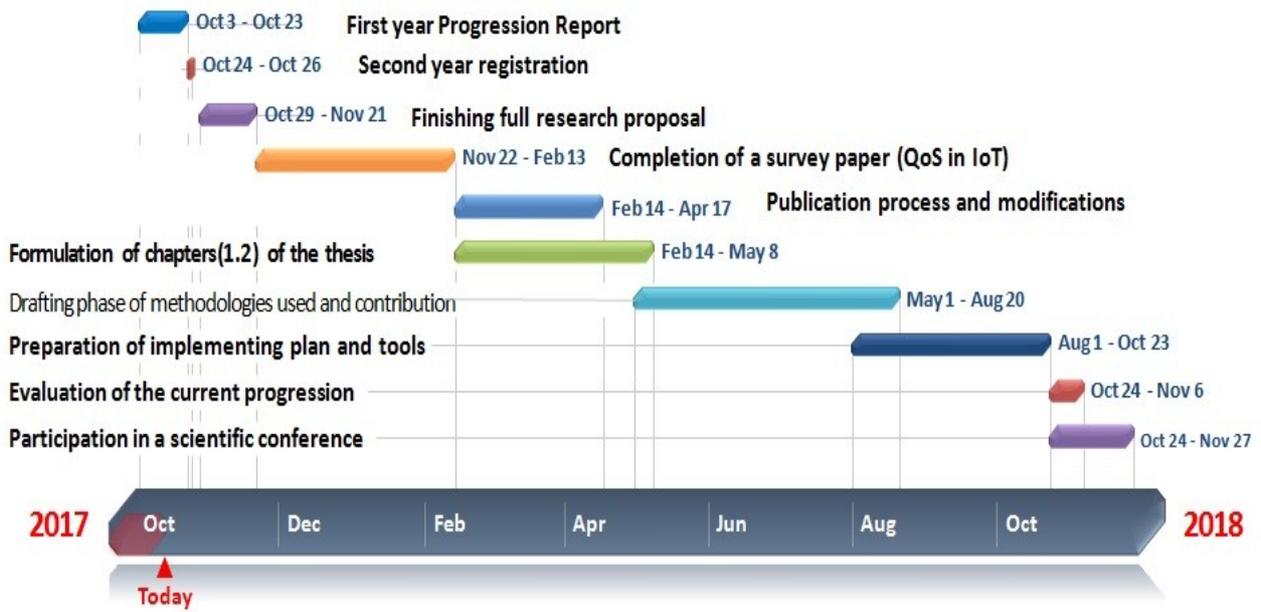


Fig:2

Third Year:

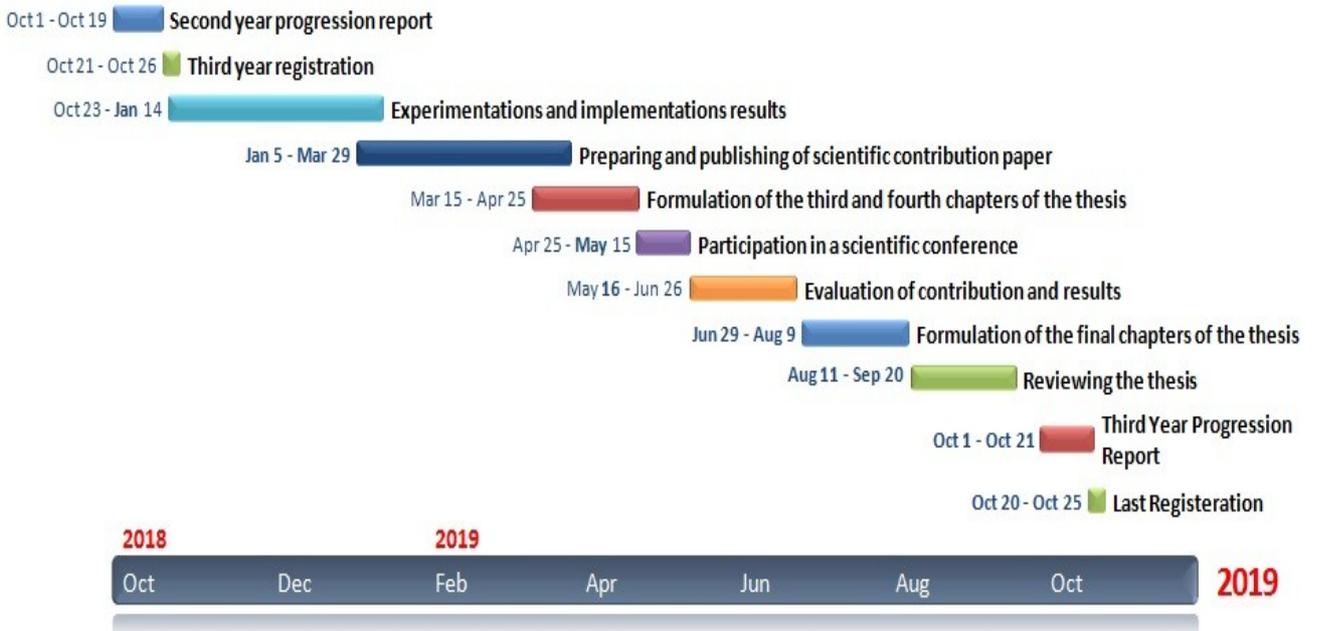


Fig:3

Quality of Service (QoS) for IoT Services

TABLE OF CONTENTS

ABSTRACT.....i

ACKNOWLEDGMENTS.....ii

TABLE OF CONTENT.....

LIST OF TABLES.....

LIST OF FIGURES

CHAPTER I: Introduction

CHAPTER II: Background and Literature Review

2.1 Internet of Things Overview

2.1.1 IoT Definitions

2.1.2 IoT Technologies

2.1.3 IoT Applications

2.1.4 IoT Communications Models

2.2 History of Internet of Things

2.2.1 IoT Historical View

2.2.2 IoT Events which have help shaping the world of IoT

2.3 IoT Emerging Economy and Development Issues

2.3.1 IoT Economic

2.3.2 IoT Development

2.1 IoT Fundamental Characteristics

2.1.1 IoT Fundamental Characteristics

2.1.2 IoT Heterogeneity

2.1.3 IoT Interoperability Standards Issues

CHAPTER III: Quality of Service Issues

3.1 Quality of Service (QoS) Background

3.1.1 QoS Definitions

3.1.2 QoS Parameters

3.1.3 QoS Metrics

3.2 QoS Management Schemes

3.2.1 QoS xxxxx

3.2.2 QoS xxxxx

3.3 Still Planning

3.3.1 QoS xxxxx

3.3.2 QoS xxxxx

CHAPTER IV: Quality of Service for IoT Services

4.1 IoT Services Requirements

4.1.1 IoT Service Features and Characteristics

4.1.2 IoT Service Specifications

4.1.3 Service Ontology and Service Description

4.2 Quality of Service for IoT Services Questions

4.2.1 Unique Quality of Services Challenges of IoT Devices

4.2.2 Aspects and Considerations

4.2.3 Problem of Heterogeneity

4.2.4 Problem of Interoperability

4.3 Quality of Service Parameters and Metrics of IoT Services

4.3.1 Identifying IoT Quality of Service Parameters

4.3.2 Identifying IoT Quality of Service Metrics

4.4 QoS Architecture Vision Overview

4.4.1 Ubiquitous Network Architecture

4.4.2 Application-Layer Overlay Network Architecture

4.4.3 Service Oriented Network Architecture

4.4.4 Others Network Architecture

4.5 IoT Quality of Service Standards and Protocols

4.5.1 Integration of Constrained Devices into the Internet

4.5.1.1 IEEE 802.15.4 Based Solutions

4.5.1.2 QoS Standards and Protocols Overview of IETF Standardization

4.5.1.3 Limitations of Assuring QoS and Problems

CHAPTER V: Related Work

CHAPTER VI: Proposed Contribution

6.1 Structure, Dimensions and Requirements

6.2 Methodologies Followed (Surveys, Measures, Data Analysis)

6.3 Technologies Used

6.4 Experimentations and Implementations Work

6.5 Evaluation and Assessment of the Proposed Solutions

CHAPTER VII: RESULTS

CHAPTER VIII: CONCLUSION

REFERENCES

Appendix A

BIBLIOGRAPHY

Big Graph Indexing and Compression

Hocine Imane

Advisors : Dr. N. Nouali-Taboudjemat and Dr. Ahcène Bendjoudi

Team : DTISI, CERIST

Abstract—There are numerous services that require a time reduction of graph queries including bioinformatics, social networking, and road networking. The best solution to get an optimal request time is by building an index in an offline phase. Traditional graph indexing techniques cannot be generalized for all graph queries, each type of queries (or two in the best case) has its specific technique. However, the fast increase of data volume (Big Data) has led to what we call Big Graphs. Thus, the volume of traditional calculated indexes of these big graphs is bigger than the volume of the graph itself. In this thesis, we propose an extension of existing indexing techniques that supports more sophisticated queries, and then we present our reduction method of space occupied by the obtained indexes, in this method, we use graph compression techniques, as well as compressed indexes techniques that have been developed on texts and trees. Finally, the proposed structures will be validated theoretically and experimentally. Experimental validation will be carried out on large graphs obtained from social networks, transport networks, and biology. And tests will be carried out on parallel architectures (cluster, multi-core CPU and GPU).

I. INTRODUCTION

Technological advances such as social networking, internet of things and cloud computing have led to a rapid increase in data volume. In fact, it is estimated that every two years the current data volume will double. Thus, Data added in petabytes, will grow to zettabytes in measure. Nowadays, data isn't only characterized with its great volume, but also with the speed of data generation (velocity), a variety of modalities and structures, variability, veracity and value of Data (6Vs). These aforementioned characteristics define what is now commonly referred to as Big Data.

In most services such as Social Networks (Facebook, Tweeter) and Google Maps (Road Networks), data is modelled by graphs. Graph Data is an important type of data storage structures that represent efficiently the relation between elements of data where nodes refer to the elements and edges to the relations. However, in the context of big data, this type of structure has evolved to become what we call big graphs. In accordance with the big data aspect, Big Graphs fit into the set of the predefined 6 Vs. It is inefficient to answer Big graph queries immediately without a pretreatment phase. This phase that consists of indexing, is indispensable to optimize the big graphs processing. Hence, we must prepare answers queries. Traditionally, we prepare only most or recent frequent queries. Indexing Graphs aims to speed up the graph query processing [5] and reduce computations complexities [4]. In this work, we will present a state of the art of indexing methods.

Finally, we aim to propose a sophisticated big graph index that will be characterized by:

- a combination of the maximum of query types.
- a reduced memory space (a compressed index).

- an optimal request time.

II. PROBLEMATIC

Traditional methods and techniques in processing data had shown their limits in Big Data field. It is clear that challenges of big data are the 6Vs in addition to the complexity of the used algorithm. The same goes for Big Graphs treatment, knowing that it generally necessitates a processing of all the nodes and edges (or maybe a large part); we are in an area of processing edges and nodes in petabytes. Thus, the storage and processing of such graphs have become convoluted. The simplest query "Reachability" is very costly with such graphs let alone the more complex ones such as Graph Matching or Partitioning. Certainly, a lot of methods and algorithms have been elaborated to manipulate graphs, but managing a continuous pruning (**Big Graphs**) remains a very challenging problem.

III. ACCOMPLISHED WORKS

In this first year of thesis, we conducted research on both big graph processing systems and indexing methods.

A. Big Graph Processing Systems

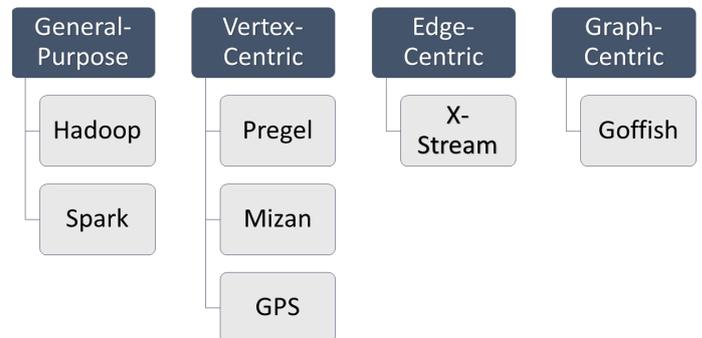


Fig. 1: Classification of distributed graph processing systems

Firstly, we carried a reasearch about large graph processing. Figure 1 summarizes the different systems studied:

a) **General purpose:** Systems for any distributed processing whether it is structured or unstructured data processing. They are not suitable for graph data processing because. They need complex MapReduce steps (chaining MapReduce jobs). Ex: Hadoop and Spark.

b) **Vertex centric:** Systems based on a simple distributed programming abstraction "think like a vertex". it is easy to programming but has often an excessive network messages.

c) *Edge centric* : A novel paradigm that push the computing from node to edge of the network. Data storage, control and trust decision are in the edge. Ex: X-Stream

d) *Graph centric* : An optimization of vertex centric systems. they address the problem of the excessive network messages by allowing messages to flow freely inside a partition of a processed graph.

In fact, several systems have been installed such as: **Hadoop**, **Spark**, **GPS** and **Giraph/Giraph++**, on the High Performance Computing Platform IBNBADIS (www.ibnbadis.cerist.dz). These installations serve to accelerates our research on indexing big graphs by preparing environments to test studied and proposed methods of big graph indexing.

B. Graph Indexing Methods

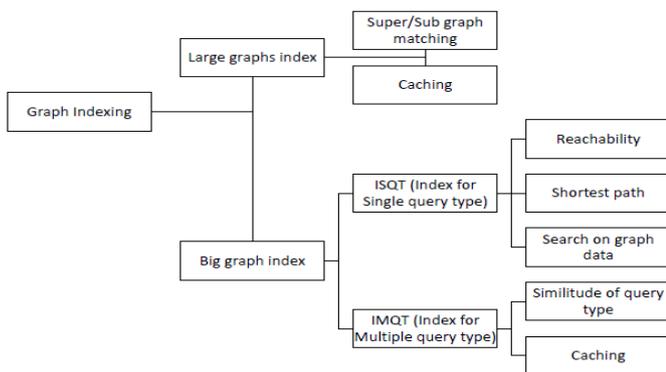


Fig. 2: Classification of Graph indexing techniques

We have developed a taxonomy of existing works that classifies techniques and methods into categories in order to have a clear vision about existing solutions of our problematic. Figure 2 summarizes the aforementioned taxonomy. Basically, we have observed that there is a variety of graph indexing methods, but some of them are specifically used for large graph databases and others for indexing a single big graph. Thus, our first dimension of this taxonomy indicates whether the indexing method is for a Big Graph or Large graphs¹. Moreover, the second dimension indicates whether the index is built to request a single graph query type or multiple types. In addition, under the dimension of multiple query types, methods can be categorized into two types according to the underlying combination's reason that can be either the resemblance in query types (ex: path queries (reachability and shortest path)) or caching (i.e. storing all recent queries regardless of their query types). Moreover, every type is by itself a field of research.

1) *Large Graphs indexing*: Many modern applications including chem-informatics, bio-informatics, and pattern recognition needed to manage large graph databases. As example: Identification of pattern or graph matching, given a Graph database of genomes of individuals (or fingerprint minutiae) $D = G_1, \dots, G_n$ and a graph G_Q . We need to match G_Q and

¹Big Graph significates a single graph with a very large number of vertices and edges, while large graphs significates numerous graphs (graph database).

identify the owner of this genome (or fingerprint minutiae). This query processing is usually known as the (sub) graph isomorphism problem [3]. Though, in reality, objects are affected by noise, so this kind of exact matching is very restrictive [7]. Thus, similarity-based search in graph databases became a basic operation in graph processing [7].

To meet the challenge of processing such large graph databases, numerous approaches of graph indexing have been proposed [6], [7], [5], [4]. We distinguish two types of large graphs indexings, indexing last queries (caching) [8] and indexing a single query type for example sub graph matching [7].

2) *Big Graph indexing*: Recently, novel services as social and road networking has produced single big graphs. These graphs have several billions of nodes and edges. At this stage, the fundamental query "Reachability" is became complex because of the necessity of traversing a very larger number of edges (edges in petabytes). Big graph indexing algorithms have been developed. These algorithms allow to accelerate the answers for certain classes of queries such as reachability [2], Shortest path [6], search on graph data, graph partitioning.

IV. CONCLUSION

This first phase of research allowed us to clearly identify the posed problem. We opted for a classification based on the type of graph queries. In addition, we carried a study about distributed large graph processing systems. We prepared a platform of tests by installing different systems in **Ibn-Badis** cluster.

V. PERSPECTIVES

Our objective now is to study graph compression methods and to search for an optimal index that combines the maximum number of queries. In addition, we aim to propose a reliable compression technique which reduces memory space to the maximum and minimize the decompression time or ideally have the possibility of getting information without decompressing the index in order to have an optimal time request.

REFERENCES

- [1] Malewicz, Grzegorz, et al. "Pregel: a system for large-scale graph processing." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.
- [2] YU, Jeffrey Xu et CHENG, Jiefeng. Graph reachability queries: A survey. In : Managing and Mining Graph Data. Springer US,p. 181-215, 2010.
- [3] H. He and A. K. Singh, "Closure-tree: an index structure for graph queries," in ICDE, pp. 38-38,2002.
- [4] Xiaoli Wang, Xiaofeng Ding, Anthony KH Tung, Shanshan Ying, and Hai Jin. An efficient graph indexing method. In Data Engineering (ICDE), 2012 IEEE 28th International Conference on, pages 210-221. IEEE, 2012.
- [5] Jing Wang, Nikos Ntarmos, and Peter Triantafillou. Indexing query graphs to speed up graph query processing. 2016.
- [6] Jouni Sirén, Niko Vaalimäki, and Veli Mäkinen. Indexing graphs for path queries with applications in genome research. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 11(2):375-388, 2014.
- [7] Wang, Xiaoli, et al. "An efficient graph indexing method." Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE, 2012.
- [8] WANG, Jing, NTARMOS, Nikos, et TRIANTAFILLOU, Peter. Indexing Query Graphs to Speed Up Graph Query Processing. 2016.

Vers Une plate-forme Open source pour le traitement automatique de la langue arabe

Taha Zerrouki

Directeurs de thèse :Amar Balla

Équipe :EIAH

Résumé— L'objectif est de concevoir une plate forme ouverte open source pour le traitement automatique de la langue arabe, c-à-d des outils simples, multi objectifs, ré-utilisables, qui s'intègrent entre eux pour construire des systèmes plus complexe et qui peuvent être re-utilisés pour d'autres objectifs, et par d'autres développeurs.

Mots- clés : traitement automatique de la langue, open source, plate-forme, langue arabe.

I. INTRODUCTION

La langue arabe est une langue très riche, en son écriture, vocabulaire, grammaire, et morphologie, ce qui rend son traitement automatique complexe, et nécessite plus d'efforts, et des méthodes plus intelligentes.

Nous pensons que le problème essentiel et crucial du traitement automatique de la langue arabe ne réside pas en sa complexité par rapport à d'autres langues, mais dans le manque d'outils et de ressources en accès libre, indispensables pour tout traitement automatique.

Notre papier est organisé comme suite : la problématique, les travaux réalisés, ensuite on décrits nos perspectives.

II. PROBLÉMATIQUE

Les ressources linguistiques basiques comme les listes de mots, les dictionnaires, les analyseurs morphologiques et syntaxiques, s'ils sont disponibles, facilitent les tâches de manipulation des données complexes, ce qui confère aux chercheurs un gain en temps et en efforts. De plus, l'aspect Open source des outils et des ressources, étend les possibilités pour le développement d'applications encore plus complexes, et plus efficaces.

Nous avons penser à concevoir et développer une plate-forme ouverte open source, qui offre des APIs pour les développeurs d'applications diverses, et aussi pour construire des systèmes complexes à base de ces outils de base.

Par exemple, le développement d'un système de synthèse vocale pour les textes arabe, devient plus simple si les textes à lire sont voyellisés, car les textes arabes sont généralement non voyellisés, ce qui rend la tâche de développement d'un synthétiseur très difficile. La disponibilité d'un voyellisateur des textes arabes est plus que nécessaire, et elle facilite donc la synthèse vocale.

D'autre part, le voyellisateur peut être utilisé à d'autres objectifs, comme l'apprentissage de la langue, l'assistance à la lecture des textes pour les journalistes et les conférenciers; et enfin l'ajustement des textes précis comme la poésie et les Hadiths du prophète.

Le voyellisateur lui-même nécessite un ensemble d'outils et ressources, tel que l'analyseur morphologique, syntaxique, les listes des mots, les dictionnaires et les corpus de textes, etc... Ces mêmes outils sont également indispensables à d'autres applications comme la correction d'orthographe entre autres.

Un outil de segmentation de textes arabes (light stemmer) pour l'indexation et la recherche des données dans des textes, et pour également la construction d'analyseurs morphologiques plus complexes.

Comme nous pouvons citer les tâches afférentes à ces outils :

- Préparer les outils de base nécessaires pour construire un future voyellisateur de textes arabes, comme un corpus de textes voyellisés et les dictionnaires des collocations.
- Collecter des corpus de textes disponibles en accès libre, afin de permettre aux développeurs d'avoir accès à ces ressources.
- Développer un dictionnaire de mots qui offre les informations et les champs spécifiques pour le l'analyse morphologique.

L'approche adoptée dans notre étude est de décomposer les tâches en des projets élémentaires pour constituer des projets basiques réutilisables pour d'autres besoins. Nous voulons faire les briques de base pour des systèmes de traitement automatique de la langue arabe, en aidant aussi les développeurs à surmonter leurs petits problèmes concernant les textes en langue arabe.

Les besoins actuels en matière d'outils du traitement automatique de la langue arabe se classent en deux catégories :

- 1- les ressources linguistiques comme les listes des mots et les corpus de textes;
- 2- et les outils de programmation que nous avons développé une gamme de produits que nous pouvons citer ci-après.

L'objectif général est de développer des API réutilisables par les développeurs, et les intégrer pour construire des systèmes plus évolués.

III. TRAVAUX RÉALISÉES

On a conçu un système de traitement automatique qui contient des outils réutilisables pour des multiple objectifs.

Le schéma globale de notre système est décrit par le schéma suivant:

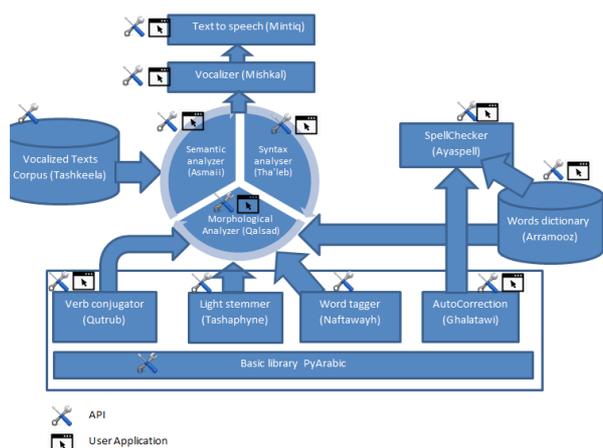


Figure 1: Structure de notre système

A. Outils

Mishkal, Arabic text vocalization : La voyellisation est une tâche ardue à développer, donc on a pensé à développer un prototype. La fonctionnalité la plus importante de notre voyellisateur est d'assister l'utilisateur pour corriger les erreurs du processus automatique [15 ; 16 ; 17 ; 18 ; 19 ; 20 ; 21 ; 22].

Outrub, arabic verbs conjugation: Une simplification de l'algorithme de la conjugaison des verbes arabes, dans les différents temps [23; 24].

Arabic SpellCheck: Le correcteur d'orthographe "Ayaspell" est un composant essentiel de notre système, car ce service est nécessaire pour la correction des erreurs lors de l'écriture [25 ; 26 ; 27 ; 28].

Adawat: Arabic text tools : Une boîte à outils pour la manipulation des textes arabes. Cet outil est une application et réutilisation des différentes fonctionnalités dans notre système.

B. Bibliothèques

Python and Arabic Pyarabic: une bibliothèque de manipulation des routines de base pour la langue arabe, tel que la suppression des voyelles, la normalisation, l'extraction des entités nommées, la manipulation des phrase de nombres, etc [12 ; 13 ; 14].

Tashaphyne : une bibliothèque de segmentation légère des mots qui permet de réaliser des analyseurs morphologiques modifiables et configurables [6;7;8;9;10;11] .

Qalsadi une bibliothèque d'analyse de texte arabe voyéllisés et semi-voyéllisés, cet analyseur est plus performant que Tashaphyne, qui est une bibliothèque de base [29;30].

Naftawayh: Arabic word tagger : cet outil permet de détecter le type du mot arabe selon les propriétés linguistiques du mots

C. Ressources Linguistiques

Tashkeela: Un corpus des textes voyéllisés qui contient 74 millions de mots [17 ; 21 ; 44 ; 45 ; 46 ; 47].

Arramooz Al-waseet: Arabic dictionary for morphology analysis : Un dictionnaire arabe orienté analyse

morphologique, il est basé sur le dictionnaire de base de la correction d'orthographe arabe (AYASPELL), réutilisé afin de promouvoir l'analyse morphologique [31 ; 32 ; 33].

AL-Radif, synonyms, antonyms, broken plural, rhymes dictionary on sinonimi : Un dictionnaire des synonymes et les antonymes arabes [34].

La liste des mots vides (Arabic stop words) : cette liste est utilisée dans l'analyse morphologique, la recherche, et l'extraction de données [36 ; 37 ; 38 ; 39 ; 40 ; 41].

Arabic Open corpora : Un répertoire des corpus d'utilisation libre, open, disponible, il recense 6 corpus des textes arabes disponibles et libres [25 ; 37 ; 42 ; 43].

Durant la cinquième année, on a travaillé sur l'implémentation des outils cités, et la préparation des articles à publier. On peut citer deux principaux travaux comme la valorisation d'une ressource linguistique des textes voyéllisés et la synthèse vocale.

La publication d'un article sur le corpus Tashkeela : le corpus des textes voyéllisés qui continent plus de 75 millions de mots, à partir de 97 livres de l'arabe classique, et de l'arabe moderne. Le corpus est disponible gratuitement sur le internet ¹.

Le papier sur ce corpus est intitulé :

"Tashkeela: Novel corpus of Arabic vocalized texts, data for auto-diacritization systems.", publié sur Data in Brief 11 (2017) [44].

Le projet de la synthèse vocale réalisé [2] en tant que projet de fin d'études d'ingénieur, par les étudiants Imed Sakraoui, Fateh Bouderdara. Le projet consiste à développer et améliorer le support de la langue arabe en utilisant un système de synthèse vocale open source dit eSpeak [3].

Un autre projet est développer pour amener le support de la langue arabe dans un environnement mobile, afin de permettre une utilisation plus large de cet outil surtout pour les aveugles et les sourds. Ce développement est réalisé par Bourabaa Mounir et Redha [4] à l'ESI.

IV. PERSPECTIVES

Comme perspectives, nous allons améliorer nos outils et ressources linguistiques afin d'obtenir des meilleurs résultats, rédiger des documents pour documenter les outils et les ressources réalisées. Décrire ces ressources pour faciliter leurs utilisations.

V. CONCLUSION

Notre travail consiste à réaliser une gamme de produits (ressources et programmes) pour le traitement automatique des textes arabes, on a réalisé un ensemble d'outils open source, disponible comme source code et comme applications téléchargeables et des ressources linguistiques libres.

Pour cette années on a travaillé sur la valorisation de nos travaux via la rédaction des papiers sur ces outils.

¹ <http://tashkeela.sourceforge.net/>

Références

- [1] "A Rule-Based Correction Method of Common Arabic Native and Non-Native Speakers' Errors", Wajdi Zaghouni, Taha Zerrouki, Amar Balla, ACL-IJCNLP, Beijing, China, July 26-31, 2015.
- [2] "Amélioration d'un outil de synthèse vocale open source pour la langue arabe.", PFE réalisé par les étudiants Imed Sakraoui, Fateh Bouderdara, ESI, 2014/2015.
- [3] <http://espeak.sourceforge.net>
- [4] « Synthèse vocale open source pour la langue arabe dans un environnement mobile », PFE réalisé par les étudiants Bourabaa Mounir Et Redha, ESI, 2015/2016.
- [5] Correcteur open source des erreurs grammaticales communes en langue arabe, PFE réalisé par les étudiantes KALI Imen, et TCHOKETCH Karima Médèa, 2015/2016.
- [6] Dahab, Mohamed Y., Al Ibrahim, and Rihab Al-Mutawa. "A comparative study on Arabic stemmers." *International Journal of Computer Applications* 125.8 (2015).
- [7] Albukhitan, Saeed, Tarek Helmy, and Ahmed Alnazer. "Arabic ontology learning using deep learning." *Proceedings of the International Conference on Web Intelligence*. ACM, 2017.
- [8] Mourad, Ahmed, Falk Scholer, and Mark Sanderson. "Language Influences on Tweeter Geolocation." *European Conference on Information Retrieval*. Springer, Cham, 2017.
- [9] Jaafar, Younes, et al. "Enhancing Arabic stemming process using resources and benchmarking tools." *Journal of King Saud University-Computer and Information Sciences* 29.2 (2017): 164-170.
- [10] Oraby, Shereen, Yasser El-Sonbaty, and Mohamad About El-Nasr. "Exploring the Effects of Word Roots for Arabic Sentiment Analysis." *IJCNLP*. 2013.
- [11] Ezzeldin, Ahmed Magdy, Yasser El-Sonbaty, and Mohamed Hamed Kholief. "Exploring the effects of root expansion, sentence splitting and ontology on arabic answer selection." *Natural Language Processing and Cognitive Science: Proceedings 2014* (2015): 273.
- [12] Al-Horaibi, Lamia, and Muhammad Badruddin Khan. "Sentiment analysis of Arabic tweets using text mining techniques." *First International Workshop on Pattern Recognition*. International Society for Optics and Photonics, 2016.
- [13] Mikhael, Kamal Abou. "The Greek-Arabic New Testament Interlinear Process: greekarabcont. org." *LRE-REL2* (2014): 1.
- [14] Abozinadah, Ehab A., and James H. Jones Jr. "IMPROVED MICRO-BLOG CLASSIFICATION FOR DETECTING ABUSIVE ARABIC TWITTER ACCOUNTS."
- [15] Azmi, Aqil M., and Reham S. Almajed. "A survey of automatic Arabic diacritization techniques." *Natural Language Engineering* 21.3 (2015): 477-495.
- [16] Bahanshal, Alia O., and Hend S. Al-Khalifa. "A first approach to the evaluation of Arabic diacritization systems." *Digital Information Management (ICDIM), 2012 Seventh International Conference on*. IEEE, 2012.
- [17] Alansary, Sameh. "Alserag: An Automatic Diacritization System for Arabic." *International Conference on Advanced Intelligent Systems and Informatics*. Springer International Publishing, 2016.
- [18] Aouichat, Asma, and Ahmed Guessoum. "Building TALAA-AFAQ, a Corpus of Arabic FActoid Question-Answers for a Question Answering System." *International Conference on Applications of Natural Language to Information Systems*. Springer, Cham, 2017.
- [19] El-Haj, Mahmoud, and Paul Edward Rayson. "OSMAN: a novel Arabic readability metric." (2016): 250-255.
- [20] Rebai, Ilyes, and Yassine BenAyed. "Arabic speech synthesis and diacritic recognition." *International Journal of Speech Technology* 19.3 (2016): 485-494.
- [21] Fashwan, Amany, and Sameh Alansary. "SHAKKIL: An Automatic Diacritization System for Modern Standard Arabic Texts." *WANLP 2017 (co-located with EACL 2017)* (2017): 84.
- [22] Alrakaf, Alaa Abdullah, and Sk Md Mizanur Rahman. "A supervised approach for word sense disambiguation based on Arabic diacritics." *Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on*. IEEE, 2016.
- [23] Imad, Zeroual, and Lakhouaja Abdelhak. "Adapting a decision tree based tagger for Arabic." *Information Technology for Organizations Development (IT4OD), 2016 International Conference on*. IEEE, 2016.
- [24] Abdelaziz, Mirad, and Bensaou Nacéra. "Al-Lamy An Automatic Conjugation System of Arabic Verbs."
- [25] Mostefa, Djamel, Omar Asbayou, and Ramzi Abbes. "TECHLIMED system description for the Shared Task on Automatic Arabic Error Correction." *ANLP 2014* (2014): 155.
- [26] Attia, Mohammed, et al. "Improved spelling error detection and correction for Arabic." *Proceedings of COLING 2012: Posters* (2012): 103-112.
- [27] Mars, Mourad. "Toward a Robust Spell Checker for Arabic Text." *International Conference on Computational Science and Its Applications*. Springer International Publishing, 2016.
- [28] Attia, Mohammed, et al. "Arabic spelling error detection and correction." *Natural Language Engineering* 22.5 (2016): 751-773.
- [29] Aly, Mohamed A., and Amir F. Atiya. "LABR: A Large Scale Arabic Book Reviews Dataset." *ACL* (2). 2013.
- [30] Nabil, Mahmoud, Mohamed Aly, and Amir Atiya. "LABR: A Large Scale Arabic Sentiment Analysis Benchmark." *arXiv preprint arXiv:1411.6718* (2014).
- [31] Raafat, Hazem, Mohamed Zahran, and Mohsen Rashwan. "A Database Combining Different Arabic Resources with Lexical and Semantic Information."
- [32] Enaanai, Adil, and Abdelaziz Sdigui DOUKKALI. "An hybrid approach to calculate relevance in the Arabic metasearch engines." *The International Journal of Science and Advanced Technology* 2.3 (2012): 127-134.
- [33] Lebboss, Georges. *Contribution à l'analyse sémantique des textes arabes*. Diss. Paris 8, 2016.
- [34] Al-Subaihin, Afnan A., and H. S. Al-Khalifa. "Al-baseet: A proposed simplification authoring tool for the arabic language." *Communications and Information Technology (ICCIT), 2011 International Conference on*. IEEE, 2011.
- [35] Raafat, Hazem, Mohamed Zahran, and Mohsen Rashwan. "A Database Combining Different Arabic Resources with Lexical and Semantic Information."
- [36] Saad, Motaz K., and Wesam Ashour. "Arabic morphological tools for text mining." *Corpora* 18 (2010): 19.
- [37] Hassan, Youssef, Mohamed Aly, and Amir Atiya. "Arabic spelling correction using supervised learning." *arXiv preprint arXiv:1409.8309* (2014).
- [38] Alghamdi, Hanan M., and Ali Selamat. "Topic detections in Arabic dark websites using improved vector space model." *Data Mining and Optimization (DMO), 2012 4th Conference on*. IEEE, 2012.
- [39] Saber, Shimaa, Ali Ahmed, and Mohy Hadhoud. "Robust metrics for evaluating arabic OCR systems." *Image Processing, Applications and Systems Conference (IPAS), 2014 First International*. IEEE, 2014.
- [40] Misbah, Ahmed M., and Ibrahim F. Imam. "Mining opinions in Arabic text using an improved "Semantic Orientation using Pointwise Mutual Information" Algorithm." *Informatics and Systems (INFOS), 2012 8th International Conference on*. IEEE, 2012.
- [41] Alghamdi, Hanan, and Ali Selamat. "The Hybrid Feature Selection k-means Method for Arabic Webpage Classification." (2014).
- [42] Al-Thubaity, Abdulmohsen O. "A 700M+ Arabic corpus: KACST Arabic corpus design and construction." *Language Resources and Evaluation* 49.3 (2015): 721-751.
- [43] Al-Sabbagh, Rania, Roxana Girju, and Jana Diesner. "Unsupervised Construction of a Lexicon and a Repository of Variation Patterns for Arabic Modal Multiword Expressions." *MWE@ EACL*. 2014.
- [44] Zerrouki, Taha, and Amar Balla. "Tashkeela: Novel corpus of Arabic vocalized texts, data for auto-diacritization systems." *Data in Brief* 11 (2017): 147-151.
- [45] Harrat, Salima, et al. "Diacritics restoration for Arabic dialect texts." *INTER_SPEECH*. 2013.
- [46] Zayyan, Ayman A., et al. "Automatic diacritics restoration for modern standard arabic text." *Computer Applications & Industrial Electronics (ISCAIE), 2016 IEEE Symposium on*. IEEE, 2016.
- [47] Hifny, Yasser. "Smoothing techniques for Arabic diacritics restoration." *12th Conference on Language Engineering*. 2012.

Approches biomimétiques parallèles et coopératives pour la résolution du problème de la coloration par listes. Application au problème d'affectation de fréquences.

Riad HADJI.

Directeurs de thèse : Karima BENATCHBA.

Équipe : Optimisation.

Résumé— Le problème d'affectation de fréquences dans les réseaux radio-mobiles modélisé en un problème de multi-coloration par listes [2] est un problème NP-difficile pour lesquels les méthodes de résolution exactes mettent beaucoup de temps à trouver une solution optimale pour des instances de grande taille, voir de taille moyenne. Dans cette thèse, nous proposons d'utiliser deux métaheuristiques biomimétiques [3] : les systèmes immunitaires [4] et les colonies d'abeilles [7] afin de résoudre ce problème. Une hybridation avec les méthodes exactes et d'autres métaheuristiques comme la recherche locale sera également utilisée. Pour réduire les temps de traitement, nous proposons d'utiliser une parallélisation [5] de ces méthodes sur des architectures parallèles comme les grappes de machines, les processeurs multi-cœur ou encore les processeurs GPU [9] (Graphic Processor Unit).

I. INTRODUCTION.

Le problème d'affectation de fréquences (FAP) [8] est l'un des problèmes combinatoires les plus pertinents qui ont fait l'objet de plusieurs travaux de recherche. Ceci est dû essentiellement aux progrès réalisés dans le domaine des communications téléphoniques sans-fil (à l'exemple des réseaux GSM) ainsi que celui des communications satellites. Ayant pour objectif d'affecter un ensemble de fréquences aux transmetteurs tout en respectant des contraintes, le FAP peut être modélisé à l'aide de graphes. Cette modélisation nous conduit à résoudre une variante du problème de coloration de graphes (PCG) : la multi-coloration par listes.

II. PROBLÉMATIQUE

Le problème de la multi-coloration par listes étant NP-difficile, la résolution de ce dernier en utilisant des méthodes exactes prend un temps de calcul considérable, surtout lorsqu'il s'agit d'instances du problème de grande taille, voire de taille moyenne. L'utilisation des méthodes approchées ou métaheuristiques constitue une alternative efficace : elles offrent des solutions de bonne qualité en un temps de calcul réduit. Parmi ces métaheuristiques, on trouve les méthodes biomimétiques inspirées de la nature, à l'exemple des systèmes immunitaires et colonies d'abeilles. Ainsi, une hybridation de ces méthodes avec des métaheuristiques ou des approches exactes permettrait de trouver un compromis entre temps de calcul et qualité de solution.

En outre, une autre technique permet de gagner en temps de calcul : la parallélisation. En effet, ces dernières années, les

méthodes de recherche parallèles ont permis d'obtenir des résultats satisfaisants.

L'objectif de notre thèse est d'expérimenter les différentes méthodes de parallélisation et de coopération de quelques métaheuristiques biomimétiques avec des méthodes exactes ou d'autres métaheuristiques et leurs implémentations sur les différentes architectures parallèles et distribuées.

III. TRAVAUX RÉALISÉS.

A. Implémentation parallèle des systèmes immunitaires artificiels avec la recherche locale pour la résolution du problème de la multicoloration par listes sous-contraintes.

Le premier travail réalisé est l'implémentation parallèle des systèmes immunitaires artificiels avec la recherche locale pour la résolution du problème de la multi-coloration par listes sous contraintes [1].

La solution que nous avons proposée se base sur la théorie de la sélection clonale des systèmes immunitaires artificiels avec les étapes suivantes : génération d'une population de solutions initiale, phase de clonage, phase de mutation et phase d'agitation auxquelles nous combinons une phase de recherche locale. Pour ce qui de la génération des solutions initiales nous avons proposé deux heuristiques MLCR (Multi coloration par Liste sous Contraintes par Ratio) et la Multi coloration par degré (MD).

L'objectif visé par MLCR est de favoriser l'obtention de solutions complètes sans notion d'optimalité sur le nombre de couleurs utilisées, son utilisation peut donc se justifier pour des instances « difficiles ». A l'opposé, la Multi coloration par degré MD se propose de donner des solutions de meilleure qualité mais peut induire un taux d'échec élevé dans la génération de solutions complètes.

Nous avons appliqué un schéma de parallélisation pour la sélection clonale qui se traduit par des exécutions concurrentes sur des sous-populations initiales de solutions, avec une stratégie de coopération.

Tous les tests ont été effectués sur une machine dotée d'un processeur Intel Core 2 Duo 2Ghz et d'une mémoire centrale de 2Go, sous un environnement Linux (Ubuntu 10.04).

Le parallélisme pour les métaheuristiques répond à deux objectifs : amélioration du temps de calcul et amélioration de la qualité des solutions. En fixant la taille de la population à 50 individus, nous comparons les résultats obtenus par un

processus améliorant 50 solutions et celui obtenu par deux processus évoluant en parallèle et avec coopération dont chacun améliore 25 individus, cette stratégie divise la *quantité* de travail et permet donc d'améliorer le temps de calcul. L'autre stratégie testée consiste à optimiser deux populations de taille première (50 solutions) en parallèle, ce qui revient à doubler la quantité de travail et permet d'explorer plus de zones de l'espace de recherche.

B. Implémentation parallèle hybride colonies d'abeilles – méthodes exactes pour la la multicoloration par listes sous-contraintes.

Dans ce travail, nous avons proposé une solution baptisée (BCO-FAP) qui est une hybridation intégrative qui consiste en l'incorporation d'une méthode exacte (EXA-SUB) basée sur la méthode du Branch&Bound [8], dans un processus de recherche basé sur les colonies d'abeilles. Cette méthode de recherche basée sur les colonies d'abeilles est inspirée de la version améliorative de la méthode connu sous le nom *Bee Colony Optimization* (BCO) [7].

Notre approche se constitue de deux phases : la phase de génération de solutions initiales et la phase d'amélioration de solutions. Dans la première phase, afin de générer des solutions initiales au problème, nous avons proposé deux heuristiques (CIM et WPD).

Par ailleurs, dans la seconde phase, nous avons mis au point deux stratégies d'amélioration : la première, étant la stratégie de base, qui s'appuie sur des méthodes de recherche locale (Hill- Climbing et RL-Perturbation) tandis que la seconde stratégie, agressive, est une extension de la première avec l'intégration d'une méthode exacte (EXA-SUB) au processus d'amélioration.

D'un autre côté, nous avons implémenté une version parallèle de notre solution : BCO-FAP-Parallel. Cette dernière, conçue pour fonctionner sur une grappe de machines (cluster) sous une architecture maître-esclave, permet de diviser la charge totale de notre algorithme sur plusieurs nœuds de calcul et ainsi gagner en temps d'exécution. Ce gain pourrait être éventuellement exploité en vue d'améliorer la qualité de la solution.

Les tests effectués ont démontré que l'incorporation de la méthode exacte dans le processus d'amélioration, sous un certain ajustement des paramètres régissant cette dernière, donne de meilleurs résultats en termes de qualité, efficacité et convergence. Par ailleurs, les tests réalisés sur la version parallèle ont montré que cette dernière peut atteindre sur un environnement multi-processeurs (plusieurs machines), une accélération de l'ordre du nombre de nœuds de calcul impliqués, tandis que les résultats sont moins satisfaisants sur un environnement multi-cœurs (plusieurs threads sur une seule machine).

C. Implémentation parallèle hybride colonies d'abeilles – méthodes exactes pour la la multicoloration par listes sous-contraintes.

Durant l'année 2016-2017, nous avons implémenté les colonies d'abeilles artificielles (ABC) [10] sur une machine

dotée d'un processus GPU (Graphic Processor Unit) [9] [11] pour résoudre le problème d'affectation de fréquences. Nous avons proposé plusieurs stratégies de parallélisation de cet algorithme de résolution.

La première consiste à paralléliser ce dernier sur GPU en lui affectant une seule colonie. Chacune de ses abeilles est alors attribuée à un seul thread qui va se charger d'améliorer sa solution indépendamment des autres. Cette stratégie a l'avantage d'être efficace et plus simple à implémenter que les deux autres.

La deuxième approche consiste à assigner plusieurs colonies au GPU qui vont alors évoluer en parallèle. L'avantage de cette dernière est qu'elle offre les meilleurs résultats en termes de qualité des solutions.

La dernière implémentation est une stratégie de parallélisation hybride permettant de profiter au maximum de la puissance du CPU ainsi que celle du GPU. Elle a donc pour avantage de permettre d'atteindre les meilleures accélérations en général.

IV. CONCLUSION.

L'objectif de cette thèse est d'expérimenter la parallélisation de métaheuristiques biomimétiques hybridées avec des méthodes exactes ou d'autres métaheuristiques pour résoudre le problème d'affectation de fréquences modélisé en un problème de multi-coloration par liste.

Notre contribution est surtout d'analyser le comportement coopérative de ces méthodes de recherche sur des architectures parallèles comme les réseaux de machines hétérogènes, les machines multi-cœur ou encore les machines GPU.

Références

- [1] R. Hadji, N. Menni, A. Meraga, and M. Bessedik, "Parallel artificial immune system for the constrained graph list multicolouring problem". *IJMHeur* 3(1): 1-20 (2014).
- [2] M.R. Garey, and D.S. Johnson, "Computers and Intractability", W.H. Freeman and Co San Francisco, California (1979).
- [3] E.G. Talbi, "Metaheuristics from Design to Implementation", 624pp, Wiley Editions (2009).
- [4] L.N. De Castro, and J. Timmis, "Artificial Immune Systems: A New Computational Approach, Springer-Verlag, London, UK (2002).
- [5] T.G. Crainic and M. Toulouse, "Parallel Strategies for Metaheuristics", in *Handbook of Metaheuristics*, F. Glover, G. Kochenberger (Eds.), Kluwer Academic Publishers, 475-513, 2002.
- [6] D. Teodorovic "Bee Colony Optimization (BCO)". *Innovations in Swarm intelligence*, 2009: 39-60.
- [7] M. Rahoual and P. Siarry, "Les réseaux informatiques : conception et optimisation". Editions Technip, 2006.
- [8] E-G. Talbi, "Metaheuristics : from design to implementation". John Wiley & Sons, 2009
- [9] A. Delévacq, "Métaheuristiques pour l'optimisation combinatoire sur processeurs graphiques (GPU)", *PhD thesis*, Université de Reims Champagne-Ardenne, 2013
- [10] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga. "A comprehensive survey : artificial bee colony (abc) algorithm and applications", *Artificial Intelligence Review*, Vol. 42, No. 1, pp. 21–57, 2014.
- [11] T. V. Luong, , N. Melab, and E. -G. Talbi, "GPU computing for parallel local search metaheuristic algorithms". In : *IEEE Transactions on Computers*, Vol. 62, No. 1 , pp. 173–185, 2013.

Hyper-Heuristiques par sélection et génération pour le problème de Flowshop

AHMED BACHA-BENBRIKA Sarra Zohra

Directeurs de thèse : Karima BENATCHBA, Fatima BENBOUZID-SI TAYEB

Équipe : Optimisation

Abstract—Nous abordons dans le présent document l'état d'avancement de notre thèse de doctorat. Dans un premier temps, nous identifions le contexte de notre travail ainsi que les motivations derrière la problématique que nous avons identifiée, ensuite nous présentons les travaux réalisés et ceux en cours et nous terminerons par mettre le point sur l'ensemble des perspectives que nous visons à atteindre l'année prochaine.

I. INTRODUCTION

L'optimisation combinatoire (OC) occupe une place très importante en informatique. Son importance se justifie d'une part par la difficulté des problèmes d'optimisation qui sont en général NP-Difficiles et d'autre part par de nombreuses applications pratiques pouvant être formulées en un problème d'OC.

La résolution de ces problèmes est assez délicate puisque le nombre fini de solutions réalisables croît généralement de façon exponentielle avec la taille du problème. Cela a poussé les chercheurs à développer de nombreuses méthodes de résolution en recherche opérationnelle (RO) et en intelligence artificielle (IA)[1]. Vu son importance, la communauté travaillant dans ce domaine est très active. Une large variété d'approches de résolution ont été étudiées, qu'elles soient heuristiques, méta-heuristiques, hybrides, parallèles, ...etc et peuvent en un temps raisonnable, trouver des solutions d'assez bonne qualité.

En dépit du succès de ces approches pour beaucoup de problèmes réels. Elles présentent des difficultés vis-à-vis leur réutilisation sur des problèmes nouvellement produits, ou même sur de nouvelles instances de problèmes déjà existants. Ces difficultés résultent principalement de l'importance du calibrage de ces méthodes qui affecte fortement leurs performances, et également le fait qu'il nécessite l'intervention des experts.

II. STRUCTURE DU DOCUMENT

À présent, nous allons définir la problématique dans la section III, puis les travaux que nous avons réalisé durant cette première année dans la section IV. Les sections V et VI présentent les perspectives prévues pour l'année prochaine et une conclusion.

III. PROBLÉMATIQUE

Pour chaque type de problème d'OC, il y a eu plusieurs méthodes de résolutions exactes ou approchées, De plus, les métaheuristiques sont devenues beaucoup plus dépendantes du problème à résoudre. Durant Les dernières années, les chercheurs ont essayé de combiner et de faire collaborer ces dernières pour trouver la meilleure méthode qui pourrait résoudre un grand nombre de problèmes. Cependant, ils ont

trouvé qu'une méthode efficace pour un problème ou une instance de problème peut être de mauvaise qualité face à une nouvelle situation. D'où, l'apparition du théorème de No Free Lunch [2] qui prouve qu'une méta-heuristique ne peut être la meilleure pour toutes les instances d'un problème donné.

Face à ces résultats, il fallait trouver une approche indépendante du problème à résoudre, ce qui est l'idée des Hyper-heuristiques. Ces dernières ont pour objectif de traiter les problèmes avec un niveau plus haut de généralité et choisir à chaque instant quelle heuristique à appliquer pour trouver une meilleure solution.

Burk [3] définit une hyper-heuristique comme une méthode de recherche qui choisit ou génère des heuristiques (ou composants d'heuristiques) simples à implémenter afin de résoudre efficacement une instance ou une classe d'instances d'un problème d'optimisation.

Une hyper-heuristique doit être en mesure de produire des solutions efficaces dans le domaine du problème sans expertise ou intervention significative des praticiens. Cette caractéristique permet d'offrir un certain degré d'abstraction et de généralité aux méthodes de recherche par rapport aux problèmes à résoudre. Ceci, dans le but de permettre aux praticiens d'investir un minimum d'efforts lors de la modélisation d'un nouveau domaine.

Notre domaine d'application est le problème de Flowshop de Permutation. Ce dernier consiste à ordonner n jobs sur m machines disposées séquentiellement avec comme fonction objectif C_{max} qui est le temps d'achèvement de tous les jobs sur toutes les machines.

IV. TRAVAUX RÉALISÉS

Avant d'entamer n'importe quel thème de recherche, il est impératif de faire un état de l'art sur le domaine en question. Dans notre cas, nous avons jugé indispensable de diviser notre recherche sur deux volets :1) Les méthodes de résolution du problème de flowshop de permutation (PFSP) et 2) Les hyper-heuristiques.

Nous avons étudié en premier les travaux pertinents sur le problème de flowshop. L'un des travaux les plus intéressants est l'état de l'art de Ruiz [4] qui a ré-implémenté les heuristiques et méta-heuristiques proposées pour le PFSP, et ceci, sous la même configuration matérielle et le

même langage de programmation, afin de comparer leurs performance. L'une des méthodes dont le principe nous a intéressé ,et que nous comptons utiliser, est "Iterated Greedy" proposée dans [5].

En deuxième lieu, nous avons dirigé notre recherche sur les hyper-heuristiques. Ces dernières ont été appliquées sur un grand nombre de problèmes comme le flowshop hybride [6], le problème d'emploi du temps [7] [8], l'ordonnancement de production avec machines parallèles [9] et le problème du MAX-SAT [9]. Burk [3] a classifié les hyper-heuristiques selon deux dimensions :

- 1) Selon la nature de recherche : il y a deux catégories: les hyper-heuristiques de sélection et les hyper-heuristiques de génération. La première catégorie consiste 'a choisir des heuristiques de bas niveau parmi un ensemble donné pour résoudre le problème. Quant 'a la deuxième catégorie, l'idée consiste 'a construire une nouvelle heuristique 'a partir des composants des heuristiques de bas niveau pour produire la solution du problème. Dans chacune de ces deux catégories majeures, l'hyperheuristique peut être constructive ou perturbatrice.
- 2) selon la nature de l'apprentissage : soit l'apprentissage n'existe pas, soit il est hors-ligne où l'apprentissage se fait avant l'exécution de l'hyper-heuristique ou bien il est en ligne et se déroule en cours d'exécution

Une hyper-heuristique est divisée en deux couches principales et une interface intermédiaire [10] :

- 1) La couche du domaine de problème qui contient les heuristiques spécifiques au problème ; dites les heuristiques de bas- niveau, les solutions, les fonctions d'évaluation et l'instance du problème.
- 2) La couche hyper-heuristique, où les heuristiques bas-niveau sont sélectionnées ou générées. L'hyper-heuristique fonctionne à un niveau d'abstraction plus élevé et n'a aucune connaissance sur le domaine. Il a seulement accès à un ensemble d'heuristiques de bas niveau qu'elle peut appeler.
- 3) L'interface entre les deux couches est appelée "barrière de domaine". Cette dernière est nécessaire pour la généralisation. De plus son rôle principal est de permettre à l'hyper-heuristique d'être appliquée sur un nouveau problème de nature différente en ne modifiant que les heuristiques de bas-niveau et en spécifiant la fonction d'évaluation.

Nous avons travaillé cette année en collaboration avec Mr Belahdji, étudiant de PFE, avec qui nous avons proposé et implémenté une hyper heuristique de génération. L'hyper-heuristique est décrite comme suit :

- la couche domaine du problème (bas niveau) est composée d'un ensemble d'opérateurs génétiques que nous avons repris et implémenté. Nous citons à titre d'exemple les opérateurs de croisement KX, PMX, SJOX, SBOX [11] ,etc.
- La couche hyper-heuristique est un algorithme génétique (AG) qui génère une population

d'algorithmes génétiques et les déroule afin de trouver une solution de bonne qualité.

Par la suite, Nous avons testé cette hyper-heuristique sur le Benchmark de Taillard [12] et celui de VRF [13] . Les tests ont été effectués sur les ordinateurs du laboratoire et un serveur Windows au niveau du CERIST. Les résultats obtenus se sont alignés avec ceux de [4] , ce qui est très satisfaisants .

V. PERSPECTIVES

A présent, nous sommes en cours de rédaction d'un article sur l'hyper-heuristique proposée. Ensuite, nous allons travailler sur la parallélisation de cette approche sur le cluster du cerist, sur laquelle nous allons approfondir les tests. De plus, nous comptons travailler sur les l'hyper-heuristiques de selection en s'inspirant de l'algorithme IG [5].

VI. CONCLUSION

A travers cet article, nous avons présenté une synthèse sur notre recherche en première année. Ainsi, nous avons fait un état de l'art et travaillé en collaboration avec Mr Belahdji sur le développement d'une hyper-heuristique de génération et nous avons obtenus des résultats très prometteurs que nous nous comptons améliorer et publier dans une conférence.

REFERENCES

- [1] S. Salhi, "Handbook of metaheuristics," *Journal of the Operational Research Society*, vol. 65, no. 2, pp. 320–320, 2014.
- [2] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [3] E. K. Burke, M. Hyde, G. Kendall, G. Ochoa, E. Özcan, and J. R. Woodward, "A classification of hyper-heuristic approaches," in *Handbook of metaheuristics*. Springer, 2010, pp. 449–468.
- [4] V. Fernandez-Viagas, R. Ruiz, and J. M. Framinan, "A new vision of approximate methods for the permutation flowshop to minimise makespan: State-of-the-art and computational evaluation," *European Journal of Operational Research*, vol. 257, no. 3, pp. 707–721, 2017.
- [5] R. Ruiz and T. Stützle, "A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem," *European Journal of Operational Research*, vol. 177, no. 3, pp. 2033–2049, 2007.
- [6] J. A. V. Rodríguez and A. Salhi, "A robust meta-hyper-heuristic approach to hybrid flow-shop scheduling," in *Evolutionary Scheduling*. Springer, 2007, pp. 125–142.
- [7] P. Cowling, G. Kendall, and E. Soubeiga, "A hyperheuristic approach to scheduling a sales summit," in *International Conference on the Practice and Theory of Automated Timetabling*. Springer, 2000, pp. 176–190.
- [8] E. K. Burke, G. Kendall, and E. Soubeiga, "A tabu-search hyperheuristic for timetabling and rostering," *Journal of heuristics*, vol. 9, no. 6, pp. 451–470, 2003.
- [9] D. Jakobović, L. Jelenković, and L. Budin, "Genetic programming heuristics for multiple machine scheduling," in *European Conference on Genetic Programming*. Springer, 2007, pp. 321–330.
- [10] J. Swan, P. De Causmaecker, S. Martin, and E. Özcan, "A re-characterization of hyper-heuristics," 2016.
- [11] R. Ruiz and C. Maroto, "A genetic algorithm for hybrid flowshops with sequence dependent setup times and machine eligibility," *European Journal of Operational Research*, vol. 169, no. 3, pp. 781–800, 2006.
- [12] E. Taillard, "Benchmarks for basic scheduling problems," *European journal of operational research*, vol. 64, no. 2, pp. 278–285, 1993.
- [13] E. Vallada, R. Ruiz, and J. M. Framinan, "New hard benchmark for flowshop scheduling problems minimising makespan," *European Journal of Operational Research*, vol. 240, no. 3, pp. 666–677, 2015.

Towards a reference ontology for information technologies

Rokia Bouzidi

Directeurs de thèse : Dr. F.Nader

Équipe : MSI

Résumé— In our research work we aim to model information technologies. This model takes the form of a reference ontology. Once realized, this reference ontology may be used to build new information systems, to evaluate existing information systems or to help decision-makers when choosing the appropriate technology from the multiple market choices.

I. INTRODUCTION

Information technologies are considered as one of the most important components of information systems. In [2] It is expected that the IT applied solutions will improve the flow of information resources (data, information and knowledge), which will result in the simplification and improvement of business management processes. However, it represents a heavyweight financial investment for enterprises when implementing information systems.

To choose the suitable technology, decision-makers need a model on which they rely when making their choices. Even when evaluating existing information systems a structured model for information technologies is needed.

Our objective towards this research work is to build a model for information technologies to accomplish the aforementioned tasks. To do so, we organized our work in three major phases. Each phase is composed of a set of steps. First, we conducted a literature review about our domain of interest. The objective behind this phase is to collect all the existing attempts to model information technologies. Then, we tried to propose a classification for information technologies as a first step in building our reference ontology. The last phase is devoted to the evaluation of the reference ontology once built.

II. PROBLEMATIC

After the literature review we have conducted, we concluded that modeling information technologies is a gap in IS research that has not been widely dealt with. Our main problematic is about the lack of such a model of information technologies. Although its importance and the multiple uses of such a model. Yet, it exists no proposition in the literature to model and organize all the technologies that exist till now.

During our research, it was clear that modeling IT as a whole is a difficult task that needs multidisciplinary teams to work on. Thus, circumscribing the domain to model was an unavoidable choice. At this stage, we talk about the second part of our problematic which is the choice of a narrower fragment of the IT domain to model.

III. ACCOMPLISHED WORK

The result of what we have been working on is presented in a paper entitled “Towards a classification of information technologies” in the 3rd ACM International Conference of Computing for Engineering and Sciences [1].

The idea of the paper is to propose a novel classification of information technologies. To do so, the paper starts with a literature review about the definition of information technologies in order to clarify ambiguity around this concept. The result of this part is a proposition of a definition of information technologies towards a meta-model of the most important concepts around IT.

We then explained the different uses of this classification. Namely, building and evaluating information systems and helping decision-makers to choose the appropriate technologies.

We have dedicate a section to the related works on which we relied to build our classification. We talk here about the work of [3], [4] and [5].

The major part of the paper is devoted to the proposed classification. We presented in this part, the resources used to build the classification as well as the classification process that goes through three steps (Pre-selection, Classification according to the IT artifact locus of impact, Classification according to the Operational/Decisional level).

Although this work tried to classify the whole technology that can be used in enterprise information systems. Yet, it does not ensure the evolution of this technology neither a broad coverage of the domain. From here comes the necessity to choose a more precise subdomain to model. In this perspective, we introduce the concept of gamification as a trending phenomenon in building information systems.

Gamification consists of using game elements in non-game contexts [6]. This concept may be used in building information systems for a better involvement of end-users. In [7], it is clearly mentioned that even though many organizations make substantial investments into projects aimed at developing new or adapting existing IS, most of the resulting IS fail to meet their goals.

Our vision is to make a backward approach for choosing the appropriate technology for information systems. By applying gamification to information systems, motivation of end-users can be used by decision-makers to choose and implement the right technology.

IV. PERSPECTIVES

Applying gamification to information system in order to build a gamification ontology is our main objective in this next phase.

To do so, we organize our work as follow:

- ✓ Dress a research methodology to cover all the knowledge in the gamification domain based on the work of [8], [9], [10], and [11].
- ✓ Extract manually all the key concepts from the existing literature. These concepts will be the reference ontology first concepts¹.
- ✓ Using alchemy platform¹ to automatically extract key concepts from the selected articles. This step is a verification of the previous step in order not to omit any important concept.
- ✓ Once the key concepts collected, the process of building our reference ontology about gamification for enterprise information systems can be started. To build our reference ontology we follow the UPON methodology. It is a novel approach to large-scale ontology building that takes advantage of the UP and the Unified Modeling Language (UML) [12].

V. CONCLUSION

In this paper we tried to make an overview of the research we are conducting. We mainly presented the objectives of our work that aims to build a reference ontology for information technologies. We explained how modeling the whole domain of information technologies is a complex task that needs a lot of resources. For this reason, we have chosen the gamification concept that can help in choosing suitable technologies for information systems and better involve end-users when using them.

Gamification is a trending phenomenon that appeared by the end of 2010. It has been used in different fields (education, business, health, tourism, finance...) and gave promising results. Since then, a great importance was given from both academic and industrial world. However, this domain seeks an organization of the existing knowledge to make the best use of it. For this, a reference ontology is needed.

Once built, this ontology may be applied to help identifying what motivates most end-users when using information systems. By this, decision-makers can make the best choice of the appropriate technologies.

References

- [1] Bouzidi, R., Nader & F., & Chalal, R. (July 2017) towards a classification of information technologies. 3rd ACM International Conference of Computing for Engineering and Sciences.
- [2] Kowalczyk, Z., & Orłowski, C. (2014). Advanced Modeling of Management Processes in Information Technology. Springer.
- [3] Prescott, M. B., & Conger, S. A. (1995). Information technology innovations: a classification by IT locus of impact and research approach. *ACM SIGMIS Database*, 26(2-3), 20-41
- [4] Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for information systems*, 12(1), 50
- [5] Nevo, S., Nevo, D., & Ein-Dor, P. (2009). Thirty years of IS research: core artifacts and academic identity
- [6] Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15). ACM.
- [7] Hsieh, J. P. A., & Wang, W. (2007). Explaining employees' extended use of complex information systems.
- [8] Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.
- [9] Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- [10] King, W. R., & He, J. (2005). Understanding the role and methods of meta-analysis in IS research. *Communications of the Association for Information Systems*, 16(1), 32.
- [11] Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.
- [12] De Nicola, A., Missikoff, M., & Navigli, R. (2009). A software engineering approach to ontology building. *Information systems*, 34(2), 258-275.

¹AlchemyLanguage is a collection of APIs that offer text analysis through natural language processing
<https://alchemy-language-demo.mybluemix.net/>

Annotation automatique d'images

Lakhdar LAIB

Directeurs de thèse : Samy AIT AOUDIA

Équipe : (Image)

Résumé— L'annotation automatique d'images consiste à associer à chaque image un groupe de mots qui décrit le contenu visuel de l'image au moyen d'un système sans aucune intervention humaine. Nous proposons un modèle génératif à l'avantage d'annoter les images sur les trois concepts sémantiques de haut niveau (Objet, Scène et Événement). Les performances de notre système sont validées sur UIUC dataset.

I. INTRODUCTION

La croissance progressive des archives de contenus visuels sur l'internet (photos ou vidéos) a produit un besoin en techniques de recherche et indexation de larges volumes d'images chaque jour. L'annotation d'images est l'un des mécanismes d'indexation d'images pour assigner un ensemble sémantique de mots-clés. Il existe différents types d'annotation d'images : l'annotation automatique, l'annotation semi-automatique, et l'annotation manuelle. L'annotation manuelle est toujours faite par l'utilisateur, elle est très précise mais en même temps très coûteuse pour un être humain. L'annotation semi-automatique consiste à créer des annotations automatiques d'images mais nécessite l'intervention de l'utilisateur. L'annotation automatique d'images [1, 2, 3, 4] est un domaine de recherche actif durant ces dernières années, spécialement avec l'apparition des grandes bases de données d'images annotées imagenet [5] places [7] places2 [8]. Dans le but d'exploiter ces larges bases de données d'images, de nouvelles techniques d'apprentissage supervisé et non supervisé sont utilisées pour générer les annotations sémantiques des images.

Ces méthodes peuvent être considérées comme un type de classification d'images multi-classes avec un très grand nombre de classes de la taille du vocabulaire utilisé. Typiquement, l'image est d'abord analysée par l'extraction de vecteurs de descripteurs, et les mots pour l'apprentissage de l'annotation sont utilisés avec des techniques d'apprentissage automatique pour pouvoir être capable d'affecter automatiquement des annotations à de nouvelles images.

Nous proposons dans le cadre de cette thèse d'étudier l'annotation automatique d'images. Nos travaux de thèse ont pour objectif l'élaboration d'un modèle d'annotation automatique d'images. Notre modèle a l'avantage d'annoter les images sur les trois concepts sémantiques de haut niveau (Objet, Scène et Événement). D'abord en appliquant sur l'image de requête deux couches de réseau de neurones artificiel. En effet, le premier pour extraire l'annotation des objets en utilisant l'architecture Alexnet sur base des images Imagenet, le deuxième pour extraire l'annotation des scènes en utilisant l'architecture Googlenet sur base des images Places. Par l'intégration de l'efficacité de générative graphique probabiliste, notre modèle permet de combiner les annotations des objets et des scènes pour générer l'annotation d'événement ainsi qu'un raffinement automatique afin d'obtenir une annotation finale de l'image.

II.1 NOTRE MODELE D'ANNOTATIONS

Notre système a l'avantage d'annoter les images sur les trois concepts sémantiques de haut niveau en utilisant Réseau neurone artificiel (CNN) [11,12,13] et les Modèles Graphiques Probabilistes (MGP).

- Développement d'un modèle d'apprentissage des *objets* (modèle **Alex** [6] avec dataset des objets **Imagenet** [5]) pour extraire l'annotation des *objets*.
- Développement d'un modèle d'apprentissage des *scènes* (modèle **Googlenet** [9] avec dataset des scènes : **Places 205** [7]) pour extraire l'annotation des *scènes*.
- Développement d'un système basé sur le Modèle Générative graphique probabiliste [14,15] avec l'objectif de combinaison sémantique (annotation des objets, annotation des scènes) avec dataset des événements : **UIUC** [10] pour obtenir une annotation d'*événement*.
- Raffinement les annotations : pour objectif d'améliorer la représentation descriptive d'image et raffiner sémantiquement les annotations non pertinentes afin d'obtenir une annotation finale de l'image (Objet, Scène et Événement).

Les réseaux de neurones convolutifs (CNN) [11,12,13] aboutissent à de très bons résultats par utilisation de réseaux pré-entraînés sur très grands datasets. Dans notre système faire rejoindre la puissance le Modèle Générative graphique probabiliste (PGM) avec CNN, nous utilisons **1213** labels distincts d'annotation :

- **1000** annotations pour l'objet, il y a **5** termes d'annotation par image.
- **205** annotations pour la scène, il y a **5** termes d'annotation par image.
- **8** annotations pour l'événement, il existe **1** terme d'annotation par image.

III. ANNOTATIONS DES OBJETS

1. PREMIERE ANNOTATIONS DES OBJETS

Pour annoter les *objets* d'images, on utilise les réseaux de neurones convolutifs (CNN) [11,12,13]

pré-entraîné (**Alex**[6] sur **Imagenetdataset**[5]) en le ré-entraînant spécifiquement (Fine Tuning) sur notre **UIUC dataset**[10].

$$\begin{aligned}
 & p(w_{test}^{(o)}, z^{(o)}, \phi^{(o)}, \psi^{(o)} | \alpha, \beta) \\
 &= \prod_{m=1}^M \prod_{n=1}^N p(w_{m,n}^{(o)}, z_{m,n}^{(o)}, \phi^{(o)}, \psi^{(o)} | \alpha, \beta) \quad (1) \\
 &= \prod_{m=1}^M p(\psi^{(o)} | \alpha) \prod_{n=1}^N p(z_{m,n}^{(o)} | \psi^{(o)}) p(w_{m,n}^{(o)} | z_{m,n}^{(o)}, \phi^{(o)})
 \end{aligned}$$

2. ANNOTATIONS DES OBJETS SUPPER CLASS

On regroupe les annotations des objets selon «latent topic» superclasse variable z ,

$$\begin{aligned}
 p(z_i^{(o)} = k | z_{-i}^{(o)}, w^{(o)}, \alpha, \beta) &= \frac{p(w^{(o)} | z^{(o)}, \beta) p(z^{(o)} | \alpha)}{p(z_{-i}^{(o)}, w_{-i}^{(o)} | \alpha, \beta) p(w_i^{(o)} | \alpha, \beta)} \\
 &= \frac{p(w^{(o)} | z^{(o)}, \beta) p(z^{(o)} | \alpha)}{p(w_{-i}^{(o)} | z_{-i}^{(o)}, \beta) p(z_{-i}^{(o)} | \alpha) p(w_i^{(o)} | \alpha, \beta)} \\
 &\propto \frac{p(w^{(o)} | z^{(o)}, \beta) p(z^{(o)} | \alpha)}{p(w_{-i}^{(o)} | z_{-i}^{(o)}, \beta) p(z_{-i}^{(o)} | \alpha)} \\
 &\propto \frac{\beta_v + n_{k,v}}{\sum_{v=1}^V \beta_v + n_{k,v}} \cdot \frac{\alpha_k + n_{m,k}}{\sum_{k=1}^K \alpha_k + n_{m,k}} \quad (2)
 \end{aligned}$$

Annotation Objet	Recurrence in UIUC
rugby ball	64
football helmet	61
shield	45
soccer ball	45
ballplayer	30
horse cart	29
Rhodesian ridgeback	25
knee pad	15
Arabian camel	10
llama	9

Tab.1. Les 10 premiers objets les plus récurrents dans évènement : POLO.

3. ANNOTATIONS DES FINALES DES OBJETS

Avec l'introduction d'une « latent topic » superclasse variable z , nous obtenons un nouveau classement pour les annotations d'objets pour

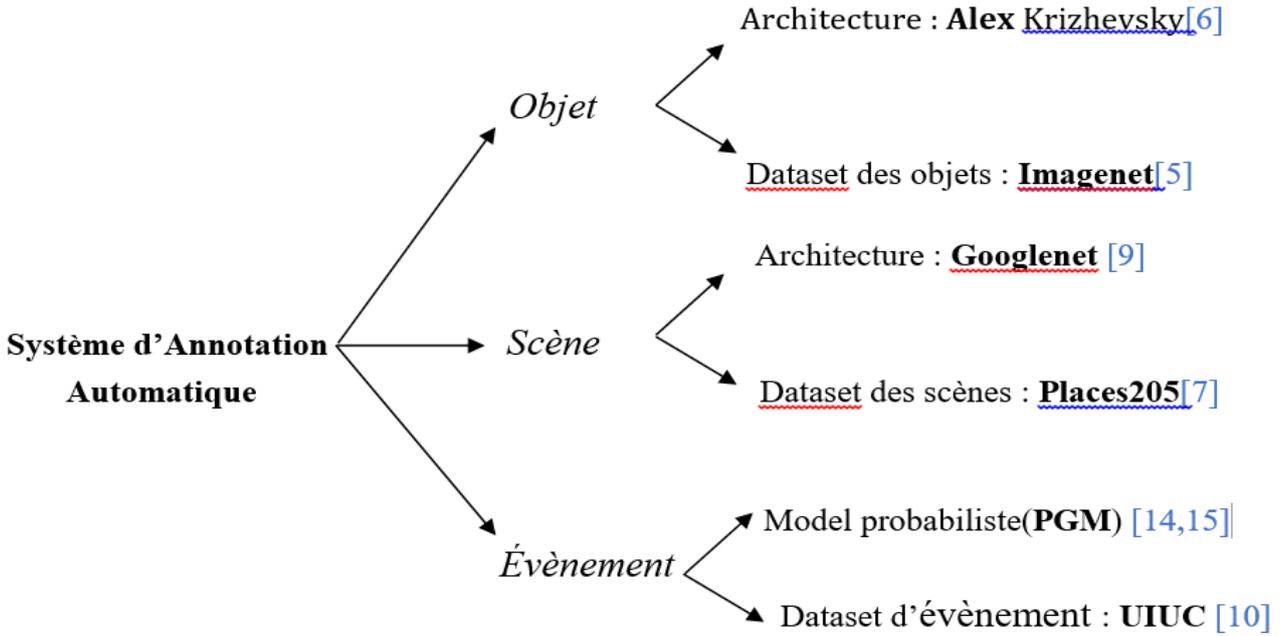


Fig.1. La structure de notre système d'annotation automatique (*Objet, Scène et Évènement*).

l'image, nous pouvons définir la probabilité conjointe d'observer un mot dans l'image par $P(w_{final}, z_k, I)$ avec la marginalisation en additionnant toutes les valeurs possibles pouvant prendre.

$$\begin{aligned}
 P(w_{final}^{(o)} | w_{Cnn}^{(o)}, I) &= \sum_{k=1}^{Na} P(w_{final}^{(o)}, z_k^{(o)}, I) \\
 &= \sum_{k=1}^{Na} P(w_{final}^{(o)} | z_k^{(o)}) P(z_k^{(o)} | I) Score_{Cnn}^{(o)}(w, I) \\
 &= Score_{Cnn}^{(o)}(w, I) \sum_{k=1}^{Na} P(w_{final}^{(o)} | z_k^{(o)}) P(z_k^{(o)} | I) \quad (3)
 \end{aligned}$$

III. ANNOTATIONS DES SCENES

1. PREMIERE ANNOTATIONS DES SCENES

Pour annoter les *Scènes* d'images, on utilise réseaux de neurones convolutifs (CNN) [11,12,13] pré-entraîné (**Googlenet**[9] sur **Places205 dataset**[8]) en le ré-entraînant spécifiquement (Fine-Tuning) sur notre **UIUC dataset**. [10]

$$\begin{aligned}
 &p(w_{test}^{(s)}, z^{(s)}, \phi^{(s)}, \psi^{(s)} | \alpha, \beta) \\
 &= \prod_{m=1}^M \prod_{n=1}^N p(w_{m,n}^{(s)}, z_{m,n}^{(s)}, \phi^{(s)}, \psi^{(s)} | \alpha, \beta) \quad (4) \\
 &= \prod_{m=1}^M p(\psi^{(s)} | \alpha) \prod_{n=1}^N p(z_{m,n}^{(s)} | \psi^{(s)}) p(w_{m,n}^{(s)} | z_{m,n}^{(s)}, \phi^{(s)})
 \end{aligned}$$

2. ANNOTATIONS DES SCENES SUPERCLASS

On regroupe les annotations des scènes selon «latent topic» superclasse variable z ,

$$\begin{aligned}
 p(z_i^{(s)} = k | z_{-i}^{(s)}, w^{(s)}, \alpha, \beta) &= \frac{p(w^{(s)} | z^{(s)}, \beta) p(z^{(s)} | \alpha)}{p(z_{-i}^{(s)}, w_{-i}^{(s)} | \alpha, \beta) p(w_i^{(s)} | \alpha, \beta)} \\
 &= \frac{p(w^{(s)} | z^{(s)}, \beta) p(z^{(s)} | \alpha)}{p(w_{-i}^{(s)} | z_{-i}^{(s)}, \beta) p(z_{-i}^{(s)} | \alpha) p(w_i^{(s)} | \alpha, \beta)} \\
 &\propto \frac{p(w^{(s)} | z^{(s)}, \beta) p(z^{(s)} | \alpha)}{p(w_{-i}^{(s)} | z_{-i}^{(s)}, \beta) p(z_{-i}^{(s)} | \alpha)} \\
 &\propto \frac{\beta_w + n_{k,w}}{\sum_{w=1}^V \beta_w + n_{k,w}} \cdot \frac{\alpha_k + n_{m,k}}{\sum_{k=1}^K \alpha_k + n_{m,k}} \quad (6)
 \end{aligned}$$

3. ANNOTATIONS DES FINALES DES SCENES

Avec l'introduction d'une «latent topic» superclasse variable z , nous obtenons un nouveau classement pour les annotations d'objets pour l'image, nous pouvons définir la probabilité conjointe d'observer un mot dans l'image par $P(w_{final}, z_k, I)$ avec la marginalisation en additionnant toutes les valeurs possibles pouvant prendre.

$$\begin{aligned}
 P(w_{final}^{(s)} | w_{Cnn}^{(s)}, I) &= \sum_{k=1}^{Na} P(w_{final}^{(s)} | z_k^{(s)}, I) \\
 &= \sum_{k=1}^{Na} P(w_{final}^{(s)} | z_k^{(s)}) P(z_k^{(s)} | I) Score_{Cnn}^{(s)}(w, I) \\
 &= Score_{Cnn}^{(s)}(w, I) \sum_{k=1}^{Nb} P(w_{final}^{(s)} | z_k^{(s)}) P(z_k^{(s)} | I) \quad (7)
 \end{aligned}$$

Annotation Scene	Recurrence in UIUC
racecourse	156
stadium/football	51
golf_course	34
fairway	23
baseball_field	22
stadium/baseball	15
train_station/platform	15
martial_arts_gym	13
pasture	12

Tab.2. Les 10 premiers scènes les plus récurrents dans évènement : POLO.

IV. ANNOTATIONS DES EVENEMENTS

Pour l'objectif de combinaison sémantique (*annotation des objets, annotation des scènes*) et obtenir une annotation d'évènement, nous utilisons le Modèle Générative graphique probabiliste (**PGM**) [14,15] avec dataset des évènements : **UIUC**. [10]

Mettre toutes les étapes ci-dessus avec l'ensemble des variables du modèle, où $X = \{e, a, t, o, s\}$, $e \in \{1, \dots, N_e\}$, $\mathbf{o} = \{\mathbf{o}_{i=1:N}\}$, $\mathbf{s} = \{\mathbf{s}_{i=1:N}\}$, $\mathbf{a} = \{\mathbf{a}_{i=1:N}\}$, $\mathbf{t} = \{\mathbf{t}_{i=1:N}\}$, l'annotation d'un évènement e peut être présentée comme suit:

$$p(X, \{\phi^{(e)}, \psi^{(e)}, \pi^{(u)}, \gamma^{(v)}\}) = p(e) \prod_{i=1}^N L_{i,o} \times L_{i,s} \quad (8)$$

où :

$p(e)$ C'est the priori probabilité de l'évènement e .

$L_{i,o}$ et $L_{i,s}$ les probabilités associées aux objets et aux scènes contiennent l'image. Ils sont donnés comme suit:

$$L_{i,o} = \prod_{m=1}^{M_i} p(a_{im} | \phi^{(e)}, e) p(o_{im} | a_{im}, \pi^{(u)}) p(r^{(o,e)} | o_{im}, \theta_{o,e}, e) \quad (9)$$

$$L_{i,s} = \prod_{l=1}^{L_i} p(t_{il} | \psi^{(e)}, e) p(s_{il} | t_{il}, \gamma^{(v)}) p(r^{(s,e)} | s_{il}, \omega_{s,e}, e) \quad (10)$$

V. DISCUSSION ET CONCLUSION

Les principales contributions de cette thèse résident dans l'amélioration de la représentation et l'extraction de contenu informationnel des images et les transformer en métadonnées textuelles sur trois concertes (Objet, Scène, Event).

- ✓ Etude les méthodes visuelles et textuelles existent sur objets, scènes et les mots clef d'annotation.
- ✓ Développement d'une méthode basée sur le deep learning (réseau de neurone) pour obtenir les annotations des objets et scènes.
- ✓ Utilisation un modèle générative pour améliorer les performances avec deux couloirs de LDA (Latent Dirichlet Allocation) le premier les annotations des objets le deuxième pour les annotations des scènes et nous avons obtenu de bons résultats.
- ✓ Evaluation notre modèle sur UIUC sport dataset avec trois niveau d'annotations (Objet, Scène, Event).
- ✓ Etude l'impact de l'utilisation des scores de confiance obtenus par annotation automatique en utilisant les mesures de performance.
- ✓ Exploiter les relations (locales/globale) et (objet/scènes) pour la reconnaissance les annotations d'image.
- ✓ Pouvoir annoter les images avec des labels à plusieurs niveaux sémantiques.
- ✓ Pouvoir annoter une image avec plusieurs labels en s'assurant qu'ils soient cohérents entre eux.
- ✓ Permettre la gestion d'un compromis entre précision sémantique des annotations et fiabilité.
- ✓ Avoir un système dont on puisse "comprendre" la décision finale (le choix des annotations). (9)

REFERENCES

- [1] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In IEEE 12th International Conference on Computer Vision, 2009.
- [2] Y. Feng and M. Lapata. Automatic image annotation using auxiliary text information. In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, 2008.
- [3] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. C. Berg, and T. L. Berg. Baby talk: Understanding and generating simple image descriptions. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.
- [4] M. Chen, A. Zheng, and K. Q. Weinberger. Fast image tagging, 2013.
- [5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. arXiv:1409.0575, 2014.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
- [7] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In NIPS, 2014.
- [8] Zhou, Bolei, et al. "Places: An image database for deep scene understanding." arXiv preprint arXiv:1610.02055 (2016).
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. CoRR, abs/1409.4842, 2014.
- [10] Li L, Li F (2007) What, where and who? classifying events by scene and object recognition. In: ICCV, pp 1{8
- [11] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In NIPS, pages 568–576, 2014.
- [12] [12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell. Caffe:Convolutional architecture for fast feature embedding. CoRR, abs/1408.5093.
- [13] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In CVPR, pages 1725–1732, 2014. 4
- [14] Lauritzen, S. L. (1996). Graphical models (Vol. 17). Clarendon Press.
- [15] Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning, 1(1–2), 1-305.
- [16] S. Gao , I. Tsang , L. Chia , Laplacian sparse coding, hypergraph laplacian sparse coding, and applications, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 35, 2013, pp. 92–104.
- [17] S. Gao , I. Tsang , Y. Ma , Learning category-specific dictionary and shared dictionary for fine-grained image classification, IEEE Trans. Image Process. 23 (2) (2014) 623–634 .
- [18] J. Wang , J. Yang , K. Yu , F. Lv , T. Huang , Y. Gong , Locality-constrained linear coding for image classification, in: Proceedings of Computer Vision and Pattern Recognition, 2010, pp. 3360–3367.
- [19] J. Yang , K. Yu , Y. Gong , T. Huang , Linear spatial pyramid matching using sparse coding for image classification, in: Proceedings of Computer Vision and Pattern Recognition, USA,,2009, pp. 1794–1801.

Modèles et outils d'annotation pour une mémoire collective entre le décideur et le veilleur dans un contexte d'intelligence économique.

Bensattalah aissa

Directeurs de thèse : Dr.Nader fahima et Pr.Chalal rachid

Équipe : (LMCS)

Résumé—Une organisation se doit d'améliorer la qualité de ses décisions par la mise en œuvre de SIADG (Systèmes Interactifs d'Aide à la Décision de Groupe). La prise de décision de groupe nécessite des mécanismes de négociation, de communication et de coordination afin d'organiser et de formaliser les interactions entre les différents acteurs impliqués. L'amélioration de la prise de décision peut porter sur une amélioration du processus cognitif de groupe. Dans le groupe se pose le problème décisionnel à résoudre, la problématique du flux communicationnel et le partage d'information. Pour assurer cette collaboration et de faciliter la communication entre les acteurs de groupe, nous proposons une mémoire collective à base des ontologies pour formaliser la communication écrite afin d'annoter les documents électroniques durant la résolution d'un problème décisionnel. L'objectif de notre travail de recherche est d'aider les décideurs de groupe grâce à l'annotation utilisée sur les documents manipulés en réutilisant le savoir-faire sur leurs activités cognitives. Pour la validation de nos propositions, nous avons retenu le processus d'intelligence économique (IE).

I. INTRODUCTION

Les décisions au sein de l'entreprise sont prises par un ou plusieurs groupes de décideurs.

Une organisation se doit d'améliorer la qualité de ses décisions par la mise en œuvre de SIADG (Systèmes Interactifs d'Aide à la Décision de Groupe). La prise de décision de groupe nécessite des mécanismes de négociation, de communication et de coordination afin d'organiser et de formaliser les interactions entre les différents acteurs impliqués. L'amélioration de la prise de décision peut porter sur une amélioration du processus cognitif de groupe.

Les acteurs impliqués dans la prise de décision coopérative utilisent des efforts mentaux et des activités cognitives considérables. Lors de ces activités cognitives, les acteurs impliqués dans la prise de décision coopérative traitent une grande masse de documents numériques. Pour faciliter leurs activités, ils utilisent différents types d'annotations sur les documents manipulés, afin d'exploiter les avantages de ces annotations.

Les acteurs impliqués dans le processus de la prise de décision analysent et traitent plusieurs documents collectés qui proviennent de sources diverses (Web, bases de données, scanners, enregistrements vidéo, rapports, rapports, etc.), avec des formats différents (Texte, image, son, vidéo, base de données, XML, etc.), avoir une sémantique différente (Différentes langues, concepts problématiques, interprétations différentes, etc.). Cette diversité pose le problème de la

connaissance non structurée ce qui est difficilement exploitable par les acteurs de groupe.

L'objectif de notre travail de recherche est d'aider les décideurs de groupe grâce à l'annotation utilisée sur les documents manipulés en réutilisant le savoir-faire sur leurs activités cognitives.

L'amélioration proposée dans le processus d'aide la décision consiste à mettre en œuvre un système de gestion d'annotation pour capitaliser et réutiliser les connaissances du groupe et de créer une mémoire collective.

Pour la validation de nos propositions, nous avons retenu le processus d'intelligence économique (IE).

II. PROBLÉMATIQUE

La prise de décision de groupe nécessite des mécanismes de négociation, de communication et de coordination afin d'organiser et de formaliser les interactions entre les différents acteurs impliqués. Notre recherche s'intéresse à la coopération cognitive dans le groupe. Les modèles de l'annotation de l'état de l'art ne permettent pas de représenter ces trois mécanismes à la fois. Le problème est donc de représenter formellement dans l'annotation ces mécanismes.

Notre première problématique est donc de modéliser l'annotation pour quelle puisse représenter les trois mécanismes d'un outil de mémoire collective.

Et la deuxième problématique est la non exploitation des annotations par les acteurs impliqués dans le processus de prise de décision, en effet les acteurs traitent d'une grande masse de ressources numériques au cours de leurs activités. Afin de faciliter leurs activités, ils utilisent différents types d'annotations sur les ressources manipulées. Les annotations permettent aux membres de groupe d'organiser, explorer, partager des idées, introduire de nouveaux concepts, créer une compréhension partagée des problèmes décisionnels, expliquer et capturer les connaissances tacites ou coopérer entre eux.

Vu que les acteurs manipulent différentes ressources hétérogènes (documents texte, vidéos, images ...) ils trouvent des difficultés pour gérer ces annotations. Pour cela il est devenu important de pouvoir annoter ces ressources hétérogènes de manière directe.

Les outils d'annotation existant ne permettent pas d'annoter des différents formats. En effet, ces outils sont souvent destinés pour annoter un seul format.

Notre deuxième objectif est de réaliser un outil d'annotation

capable de gérer, d'exploiter et de capitaliser toutes les annotations nécessaire pour les acteurs manipulés sur les différentes ressources hétérogènes.

III. TRAVAUX RÉALISÉS

Nous avons proposé une mémoire collective à base d'annotation sémantique entre les acteurs d'intelligence économique [1]. Cette proposition pour l'objectif est de structurer la communication entre les acteurs d'IE durant la collaboration pour résoudre un problème décisionnel et ainsi de capitaliser les connaissances sur leurs efforts mentaux utilisés.

Le deuxième travail réalisé est de proposer une architecture adaptative pour une mémoire collective [2]. La figure 1 illustre le modèle d'annotation pour l'aide à la décision de groupe et la figure 2 présente l'architecture adaptative pour une mémoire collective.

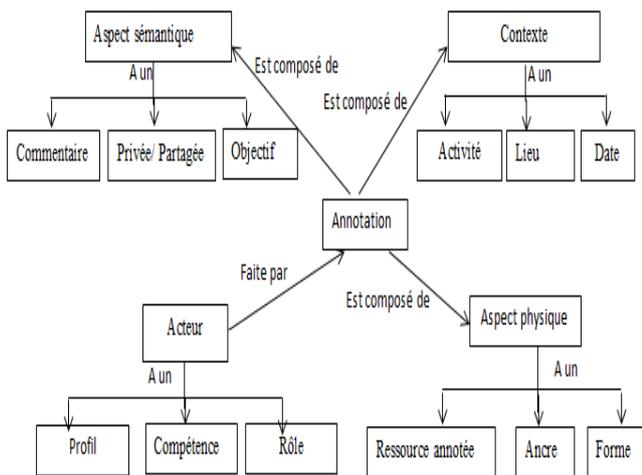


Fig.1. Modèle d'annotation pour l'aide à la décision de groupe

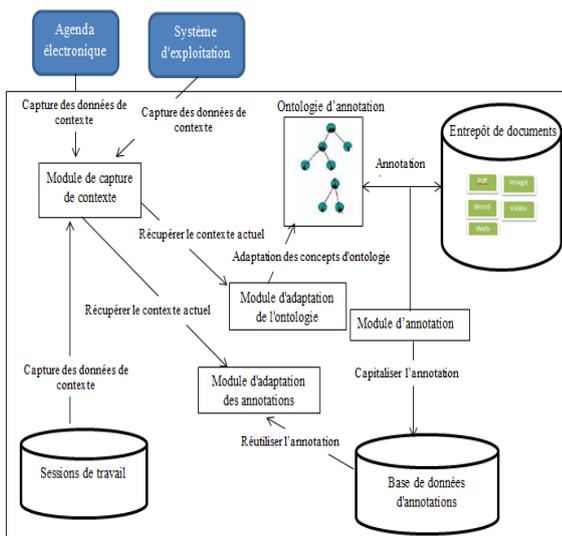


Fig.2. Une architecture adaptative pour une mémoire collective

Après l'acceptation de l'article [2] nous avons commencé la rédaction de la thèse, cette thèse est structurée en cinq chapitres. L'état d'avancement de la rédaction est de 95%.

IV. PERSPECTIVES

Nous allons structurer la thèse et nous allons préparer la présentation powerpoint de la soutenance.

Nous allons continuer travailler sur les modèles d'annotations pour améliorer nos contributions. Nous souhaitons intégrer le développement de l'interface avec les différents outils informatiques par l'utilisation des services web et ceci afin de pouvoir récupérer les activités courantes des acteurs impliqués dans le processus de prise de décision. Et ainsi nous allons étudier l'utilisation de notre application en contexte ubiquitaire. En effet, lors des déplacements pour des réunions avec d'autre acteur ou pour des études sur le terrain, il peut s'avérer très utile pour les membres de l'équipe d'offre de prendre des notes dans un contexte de mobilité.

V. CONCLUSION

Pour structure la communication et de capitaliser l'intelligence collective entre les acteurs d'IE nous avons proposé une ontologie de communication intégrée dans un outil d'annotation [1], le problème de cet outil est d'annoter un seul format (document Web). Pour résoudre le problème de l'hétérogénéité des formats, nous avons proposé une architecture adaptative pour une mémoire collective à base d'annotation sémantique [2]. Pour la validation de nos propositions, nous avons retenu le processus d'intelligence économique (IE).

Nous avons rédigé notre thèse de doctorat. L'état d'avancement de la rédaction est de 95%.

Référence

- [1] A Bensattalah, F Nader, R Chalal. "Collective memory based on semantic annotation among economic intelligence actors." International Journal of Collaborative Intelligence 2014.
- [2] A Bensattalah, F Azouaou, F Nader, R Chalal "EI-Annotate: An adaptive collective memory based on annotation ontology and context for decision making in economic intelligence" International Journal of Business Information Systems 2017 in press.

Infrastructure de communication sans fil avec qualité de service pour la gestion de crise et catastrophe

Abdelbaset KABOU

Directeurs de thèse : Mme Nadia NOUALI-TABOUDJEMAT, Mr Omar NOUALI

Équipe : Division Théorie et Ingénierie des Systèmes Informatiques - CERIST

Résumé—L'accent sera mis, au cours de ces deux pages, sur l'état d'avancement de notre thèse, en résumant la problématique, les travaux réalisés ainsi que les perspectives envisagés.

I. PROBLEMATIQUE

a) Thématique

Le sujet de notre thèse consiste à proposer une architecture de communication, répondant aux exigences relatives aux situations de catastrophes. On s'est intéressé aux Infrastructures de communication sans fil qui répondent bien, de part leur mode de fonctionnement et de configuration, à l'impératif d'un déploiement rapide et à moindre coût. Une telle infrastructure doit faciliter la collaboration entre les différentes parties intervenantes avec la qualité de service (QoS) requise pour supporter les différents services et applications (multimédia) qui seront offerts..

b) Problématique de la QoS

Un intérêt particulier est porté dans notre sujet à la problématique du maintien de la QoS pour les applications multimédia. Ce type d'applications est très sensible aux variations des conditions régissant le réseau. Des métriques comme la bande passante, la latence, la gigue etc., sont de très grande importance pour ces applications et plus particulièrement durant une situation d'urgence.

II. TRAVAUX RÉALISÉS

A. Terminologie et Concepts

Notre premier objectif était de faire un état de l'art sur le domaine de la gestion de catastrophes. Le but est de se familiariser avec la terminologie utilisée dans ce domaine et ainsi de bien distinguer entre différents concepts clefs tels que: Aléa (Hazard), Enjeu, Vulnérabilité, Risque majeur ou Catastrophe.

B. Information, communication et interopérabilité durant l'Intervention

L'objectif suivant était de s'intéresser à la communication, étant un élément indisponible pour mener à bien une coordination bien réussie.

Malgré des investissements massifs pour implémenter des nouvelles solutions de communication pour la gestion de catastrophes, leurs utilisation n'a pas réussis à faciliter le développement d'une vision/compréhension (awareness) [1]

commune lors des opérations de gestion de crises et de catastrophes. La littérature est riche des travaux qui s'intéressaient aux causes derrière ce dysfonctionnement. Dans la plupart, le manque d'interopérabilité est mis en avant comme étant un des défauts majeurs. La nécessité de revoir les aspects infrastructure de communication devient ainsi un besoins urgent.

C. Hétérogénéité et infrastructure de communication durant une opération d'Intervention

Adapter la configuration des équipements du réseau dynamiquement et offrir des interfaces de communication standards est un challenge indispensable à relever. Afin de remédier à ce problème, et avant de définir les technologies adéquates, spécifier l'ensemble des exigences que cette infrastructure doit respecter, de manière claires, était pour nous une étape indispensable. Dans ce contexte, les efforts de standardisations du domaine nous ont été très utiles notamment avec l'ensemble de besoins et d'exigences opérationnels et fonctionnels cités.

C. Proposer une architecture modulaire : la solution pour assurer l'interopérabilité

L'objectif suivant, après la mise en œuvre des besoins opérationnels et fonctionnels, était d'étudier les différentes solutions proposées dans la littérature et qui, en plus de garantir la possibilité de communiquer et de coopérer effectivement durant des situations de crise, devraient respecter au maximum les exigences discutées auparavant.

L'état de l'art nous a permis d'aboutir à une architecture abstraite qui mets en place un inter-réseau composé de plusieurs niveaux. Ces réseaux vont être organisés en couches (Fig. 1) définies à partir des fonctions fournies (i.e. Trois niveaux distincts: Terminaux mobiles, Distribution et Proxy). **La proposition a fait l'objet d'un papier soumis accepté en une conférence IEEE.**

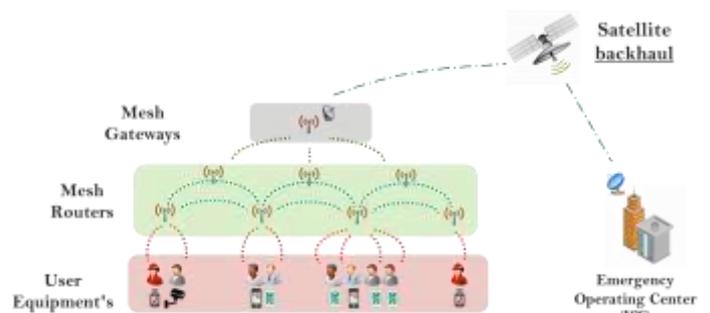


Figure 1: Architecture proposée [1]

D. Opter pour les réseaux Mesh : le meilleur choix au niveau Distribution

Une étude bibliographique nous a permis d'opter pour les réseaux Mesh, étant la solution la plus adaptée à déployer au deuxième niveau de notre architecture (niveau distribution). Cela est justifié par les options qu'offre ce type de technologie [1] notamment en termes d'auto-organisation, rapidité de déploiement, résilience, et la capacité à supporter différents types de flux.

Néanmoins, ce type de réseau est encore sujet à de moult critiques. Une parmi les plus importantes, surtout dans le contexte d'une opération d'intervention, est la rapide dégradation de débit avec l'augmentation du nombre de flux [2]. Cette limitation rend ce type de réseau très instable à partir d'un certain nombre de nœuds formant le réseau.

E. Les algorithmes Backpressure: assurer l'optimalité et la stabilité durant le routage

Backpressure est une famille d'algorithmes, immergés récemment, qui offrent des résultats très satisfaisants en termes de débit et de stabilité dans les réseaux Mesh [2].

Néanmoins, de nombreuses lacunes sont encore présentes, empêchant une large adoption de ces algorithmes. Une des limitations majeures de ces algorithmes est leurs faibles performances en termes de délai de bout-en-bout (end-to-end Delay). **Cette problématique fait l'objet de d'autres contributions [2][3]** où on propose une variante de cet algorithme offrant un nouveau mécanisme d'ajustement et de contrôle de la Qualité de service pour chaque flux circulant dans le réseau (de type UDP [2] ou TCP [3]).

La proposition non seulement offre une meilleure optimalité, stabilité, et efficacité, mais aussi implémente une différenciation en termes de qualité de services. L'idée est à la base de deux contributions :

1) Offrir une QoS ajustable suivant les contraintes de chaque flux

L'objectif de la contribution est de proposer un nouveau protocole de routage, une variante de la famille Backpressure, permettant le respect des contraintes temporelles de chaque flux. Baptisée **Lifetime-aware Backpressure** et implémentée et évaluée sous le simulateur

ns3, **la proposition fait l'objet d'un papier soumis à un journal IEEE [2] et est en deuxième tour de révision].**

2) Améliorer l'interaction entre le Backpressure et TCP

Afin d'assurer une bonne prise en charge de la Qualité de Service, la couche Transport est aussi un niveau important à ne pas négliger. TCP (Transmission Control Protocol), considéré comme le principal protocole à ce niveau, est connu pour son incompatibilité avec les algorithmes inspirés du Backpressure.

Notre deuxième contribution [2] vise à remédier à cette problématique en proposant deux variantes du Backpressure capables d'interagir efficacement avec TCP. Les deux variantes baptisées *V-Based Differentiation Backpressure* et *Queue-based differentiation Backpressure* ont été implémentées sur le simulateur réseaux ns3. L'évaluation de performance confirme la supériorité des approches proposées par rapport à d'autres algorithmes issues de la littérature. **Le travail a été soumis à une conférence IEEE [3].**

III. CONCLUSION ET PERSPECTIVES

Après une première proposition de l'architecture global de notre solution (soumise à une conférence IEEE [1]), ainsi que nos deux autres contributions (l'une acceptée à une conférence IEEE [3] et l'autre soumise pour une publication en un journal IEEE [2]), nous comptons proposer une extension qui englobe les trois contributions en un seule Framework générique pour la gestion de crises et catastrophes. En parallèle le processus de rédaction est en cours et nous comptons le finaliser d'ici novembre prochain.

References

- [1] Abdelbaset Kabou, Nadia Nouali-Taboudjemat, Omar Nouali. Toward a new Backpressure-based framework to Enhance Situational Awareness in Disaster Response. Accepted at the 4th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). IEEE, 2017.
- [2] Abdelbaset Kabou, Nadia Nouali-Taboudjemat, Soufiene Djahel, Saïd Yahiaoui, Omar Nouali Silvestro. Lifetime-aware Backpressure - a new delay-enhanced Backpressure-based routing protocol. Submitted at the IEEE Systems Journal. (Accepted with major revisions)
- [3] Abdelbaset Kabou, José Núñez-Martínez, Nadia Nouali-Taboudjemat, Omar Nouali, and Josep Mangués-Bafalluy. Analyzing backpressure routing interaction with tcp for wireless mesh networks. In International Conference on Information Networking (ICOIN), 2017. IEEE, 2017.

Critères de Performance et indicateurs pour évaluer le niveau d'apprentissage chez les apprenants dans un EIAH.

BELGROUNE BRAHIM

Directeur de la thèse : Dr.ADMANE Lotfi.

Equipe : EIAH

Résumé : notre objectif est l'évaluation du niveau d'apprentissage des apprenants dans un EIAH (Environnement Informatique d'Apprentissage Humain). Notre méthodologie est basée sur l'analyse des traces d'utilisation d'EIAH. On utilise ces traces pour calculer des indicateurs numériques. Ces indicateurs représentent un ensemble des critères d'évaluation. La représentation numérique permet d'analyser, de comparer et de classer ces critères. La classification permet, aux observateurs, d'évaluer le niveau d'apprentissage des apprenants. Dans cet article, nous présenterons l'état d'avancement des travaux sur ce sujet.

I. Introduction

Dans notre travail, l'évaluation du niveau d'apprentissage dans un EIAH revient à collecter les traces d'utilisation et les analyser pour calculer des indicateurs numériques. L'évaluation se fait, ensuite, à l'aide d'un ensemble de **critères d'évaluation** représentés par ces indicateurs.

L'utilisation des **traces d'interaction** pour évaluer le comportement des apprenants nous guide à discuter sur les systèmes à base de traces (SBT). Particulièrement, les SBT calculant des indicateurs sont des TBS-IC (Trace Based System Indicator Computation).

Dans cet article, nous exposons une méthodologie d'évaluation basée sur l'analyse de comportement des apprenants.

II. Problématique

Dans l'état de l'art, nous avons analysé quelques études descriptives et comparatives tel que l'étude effectuée par Nour EL MAWAS et Lahcen Oubahssi dans [1] sur sept plates formes et l'étude effectuée par OUBAHSSI L, GRANDBASTIEN M et CLAËS G dans [2] sur cinq plates formes. Ces études nous mènent à constater qu'aucune plate forme ne couvre complètement le cycle complet d'apprentissage. Chacune d'elles apporte des fonctionnalités correspondantes aux besoins pour lesquels elle a été conçue. La phase d'évaluation du niveau d'apprentissage est moins traitée dans la plus part des plates formes.

Ceci nous amène à nous intéresser à l'évaluation du niveau d'apprentissage en analysant les traces d'utilisation de la plateforme.

III. Travaux réalisés

A. L'état de l'art

Dans l'état de l'art, nous avons rédigé trois chapitres :

Le premier chapitre présente les EIAH en général. Il commence par un peu d'historique sur l'utilisation de l'ordinateur dans l'éducation, puis il décrit les différents types des EIAH, et enfin, il résume les différents caractéristiques de chaque type [3].

Le deuxième chapitre a été réservé pour les systèmes à base de traces. Il présente le fonctionnement de ces systèmes, leurs paramètres et leurs utilisations. Il présente aussi les différentes méthodes d'analyse des traces, de calcul des indicateurs et d'utilisation de ces indicateurs dans les différents domaines [4][5].

Le troisième chapitre a discuté sur les différentes méthodes d'évaluation des systèmes EIAH en spécifiant les exceptions de ses environnements et leurs spécifications et comment adapter ces méthodes générales pour les appliquer sur les EIAH [6].

B. Contribution théorique

Dans ce chapitre, nous avons présenté le processus d'analyse des traces d'utilisation d'un EIAH. Nous avons présenté aussi les différents modèles de données (modèle de trace, modèle d'indicateur et modèle de critère) utilisés pour modéliser les traces récupérées de la base de traces d'EIAH.

La Figure 1 représente l'architecture globale du processus d'analyse des traces pour évaluer le processus d'apprentissage.

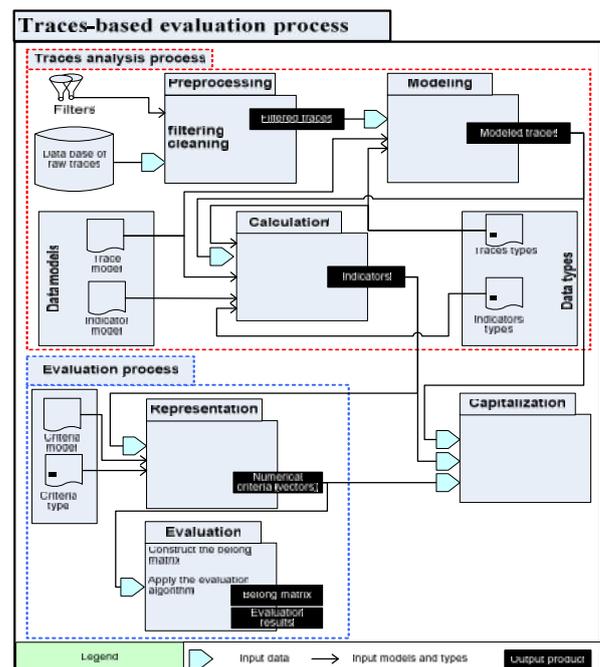


Fig. 1. Architecture du processus d'analyse des traces.

Le processus se divise en deux sous processus : le processus d'analyse de traces et le processus d'évaluation. Le rôle du premier processus est d'analyser les traces d'utilisation pour calculer des indicateurs. Cette opération se déroule sur trois étapes : (1) le prétraitement qui consiste à filtrer et nettoyer les traces brutes. (2) la modélisation qui consiste à récrire ces traces selon un modèle de trace unique. (3) le calcul qui consiste à calculer les valeurs des indicateurs. Le rôle du deuxième processus (évaluation) est d'utiliser ces indicateurs pour évaluer le comportement des apprenants à travers deux opérations : (1) la représentation des critères par les indicateurs calculés par le

premier processus. (2) l'évaluation qui consiste à appliquer l'algorithme d'évaluation sur l'ensemble des critères pour classer les apprenants (ou les groupes d'apprenants).

C. Contribution pratique

Dans ce chapitre, nous avons présenté l'architecture de la base de traces du MOODLE et l'architecture de la base des traces modélisées issue des modèles de données représentés dans le chapitre précédent. En suite, nous avons présenté l'ensemble des critères choisis pour l'évaluation du processus d'apprentissage. Pour chaque critère, on a présenté l'ensemble de traces à collecter, l'ensemble d'indicateurs représentant ce critère et la méthode à suivre pour calculer leur vecteur d'appartenance.

Selon notre méthode, chaque critère a été représenté par un vecteur de cinq éléments numériques. Chaque élément est une probabilité d'appartenance à l'une des classes : très bonne, bonne, médiocre, faible et très faible. A la fin, l'ensemble de critères se représenté par une matrice d'appartenance dont le nombre des colonnes est cinq et le nombre des lignes est le nombre des critères proposés.

La matrice (1) représente un exemple qui explique comment cette matrice représente le système d'évaluation numériquement.

$$M_{Eval} = \begin{pmatrix} V_1 \\ V_2 \\ \dots \\ V_n \end{pmatrix} = \begin{bmatrix} P_{11} & P_{21} & P_{31} & P_{41} & P_{51} \\ P_{12} & P_{22} & P_{32} & P_{42} & P_{52} \\ \dots & \dots & \dots & \dots & \dots \\ P_{1i} & P_{2i} & P_{3i} & P_{4i} & P_{5i} \\ \dots & \dots & \dots & \dots & \dots \\ P_{1n} & P_{2n} & P_{3n} & P_{4n} & P_{5n} \end{bmatrix} \quad (1)$$

Où : $V_i (1 \leq i \leq n)$: est le vecteur d'appartenance du critère Cr_i

n : est le nombre de critères proposés par le concepteur.

P_{ij} : est la probabilité que le critère i appartient à la classe j ($1 \leq i \leq n; 1 \leq j \leq 5$).

Le tableau 1 représente l'ensemble des critères proposés pour l'évaluation de la plateforme MOODLE.

Communication/Collaboration	
Email	Number of mails (sent or received) per learner. Documents (sent or received) per learner.
Forum & chat	Number of subjects that created peer learner. Number of comments per subject and per learner. Number of learners per subject.
Wiki	Number of subjects (created or modified) per learner. Number of learners per Wiki.
Learners connectivity	Number of connections (sent or received) inter group. Number of connections (sent or received) between groups.
Accessibility/effectiveness	
Learning access	Number of login per learner. Time of login per learner.
Pages access	Number of access per page. Number of learners per page. Time of access per page.
Document	Number of documents (upload or download) per learner. Number of downloads per document.
Knowledge	
QCM & QUIZ	Number of QCM (QUIZ) per learner. Number of learners per QCM (QUIZ). Time of answer per learner. Percent of true answers per learner.
Page reading	Number of consultation of page. Percent of consultation per learner. Time of consultation per learner.
FAQ	Number of questions per learner. Number of answers per learner.

Tableau 1. ensemble des critères d'évaluation utilisé.

IV. Perspectives

Il nous reste, pour l'année prochaine, à implémenter notre proposition et l'appliquer sur la plateforme MOODLE. Enfin, l'évaluation des résultats obtenus est le dernier travail à accomplir.

La publication des résultats obtenus reste l'objectif major de cette année. Actuellement, tous les efforts se focalisent sur la l'intégration de l'algorithme d'évaluation dans la platform MOODLE .

V. Conclusion

Ce document concerne l'avancement de travail sur le sujet d'évaluation du processus d'apprentissage dans un EIAH.

Un état de l'art a été traité sur trois chapitres : le premier a présenté les EIAH, leurs caractéristiques et leurs différents types. Le deuxième a traité les systèmes à base de trace, leur fonctionnement, leurs processus et leurs méthodes d'analyse des traces. Le troisième a traité la problématique d'évaluation des EIAH en parlant sur les différentes méthodes et approches utilisées. Il traite aussi les spécifications des EIAH et la possibilité d'adapter quelques méthodes générales à ce type spécifique des systèmes.

Dans la contribution théorique, une méthodologie complète d'évaluation à base de traces a été présentée. On a présenté aussi les modèles des données utilisées pour modéliser les traces brutes d'EIAH.

Dans la contribution pratique, nous avons présenté l'architecture de la base de traces brute d'EIAH et de la base de traces modélisées. Nous avons proposé un ensemble des critères d'évaluation. Pour chaque critère, on a présenté les indicateurs qui le présentent et la méthode de calcul de leur vecteur d'appartenance.

Pour finir ce travail, il nous reste à implémenter la méthode proposée et l'appliquer sur la plate forme MOODLE pour analyser les résultats, les évaluer et les publier.

Références

- [1] Nour, EL MAWAS et Lahcen, Oubahssi. Étude comparative de plates formes de formation à distance. s.l. : GraphiT, Réf: GRAPHIT-D2.5&2.2 , 2014.
- [2] OUBAHSSI, L, GRANDBASTIEN, M et CLAËS, G. Analyse Synthétique de cinq plates-formes de formation ouverte et à distance. s.l. : (rapport interne), 2003.
- [3] Tchounikine, Pière. Précis de recherche en ingénierie des EIAH. s.l. : le net sur le Web, 2009.
- [4] Ardito, C, et al. An approach to usability evaluation of e-learning applications. s.l. : Published online: 8 December 2005, Springer-Verlag 2005, 2005. Univ Access Inf Soc (2006) 4: 270–283 DOI 10.1007/s10209-005-0008-6.
- [5] Parlangei, O, Marchigiani, E et Bagnara, S. Multimedia system in distance education: e ects on usability. s.l. : Interacting Comput 12:37–49, 1999.
- [6] SETTOUTI, Lotfi Sofiane. Systèmes à Base de Traces Modélisées :Modèles et Langages pour l'exploitation des traces d'Interactions. s.l. : thèse de doctorat, école doctorale informatique et mathématique de lyon, 2011.

Des méthodes biomimétiques pour les Problèmes d'Affectation de Fréquences (PAF)

Aicha Mahmoudi

Benatchba Karima :

Équipe : Optimisation

Année de la première inscription : 2012

Résumé— Le Problème d'affectation des fréquences est un problème d'optimisation combinatoire NP-difficile.

Les enjeux de la résolution de ce problème sont très importants, sur le plan scientifique pour produire du savoir algorithmique, et sur le plan économique pour améliorer la performance des réseaux de télécommunication.

Cette thèse présente et étudie quelques approches de résolution des PAFs basées sur des méthodes biomimétiques et sur des modèles de théorie des jeux.

Mot clés : Problème d'affectation de fréquences, Algorithmes mimétiques, Théorie des jeux, Problèmes NO-Difficile, Méta-heuristiques.

I. INTRODUCTION

Le Problème d'affectation des fréquences (PAF) dans les réseaux de télécommunication est un problème d'optimisation combinatoire NP-difficile.

Les méta-heuristiques et les méthodes biomimétiques offrent une alternative prometteuse à la résolution des problèmes d'optimisation difficiles.

Les enjeux de la résolution de ces problèmes sont très importants, sur le plan scientifique pour produire du savoir algorithmique, et sur le plan économique pour améliorer la performance de systèmes de toutes natures.

La théorie des jeux a été utilisée dans divers domaines dont l'économie, l'optimisation et autres, et a prouvé sa robustesse pour ces applications.

Cette thèse présente et étudie quelques approches méta-heuristiques pour les PAF. Elle propose des approches (hybrides ou non hybrides) de résolution des problèmes d'affectation de fréquences basées sur des méthodes biomimétiques.

II. PROBLEMATIQUE

Brièvement, un problème d'affectation de fréquence (PAF) consiste à assigner une ou plusieurs fréquences à des transmetteurs ou terminaux de sorte à optimiser un ou plusieurs objectifs et de satisfaire une ou plusieurs contraintes.

Une instance d'un PAF est un tuple (ξ, F, D, R, I) où

- $\xi = \{e_1, e_2, \dots, e_n\}$ est l'ensemble de transmetteurs.

- $F = \{f_1, f_2, \dots, f_n\}$ est l'ensemble des fréquences existantes.

- $D = \xi \times \xi \rightarrow R$, est la fonction telle que $D(e, e')$ est la distance entre les transmetteurs e et e' .

- $R : \xi \rightarrow N$, est la fonction telle que $R(e_i)$ est le nombre de fréquences demandées par l'émetteur e_i .

- $I : R \rightarrow N$, est la fonction telle que $I(d)$ est la séparation de fréquences nécessaire pour éliminer les interférences entre deux émetteurs séparés par une distance d .

Ainsi, une solution pour une instance du PAF est un vecteur α

$\alpha = \langle \alpha_1, \dots, \alpha_n \rangle$ dont α_i est l'ensemble des fréquences affectées à l'émetteur e_i , telque $|\alpha_i| = r(e_i)$.

Les contraintes à satisfaire par une solution α sont les suivantes :

i. Contraintes co-site: Un écart doit être respecté entre les fréquences attribuées aux terminaux sur le même site.

ii. Contraintes Co-station (ou du Canal adjacent): Un émetteur associé à un premier terminal perturbe la réception d'un autre terminal d'un champ adjacent, la valeur de l'écart est calculée pour chaque occurrence et dépend entre autres des caractéristiques des matériels et de leur implantation géographique respective.

iii. Contraintes duplex : Il s'agit d'un écart à respecter entre les fréquences attribuées à deux terminaux issus d'une même liaison.

III. TRAVAUX RÉALISÉS

Nous proposons dans notre thèse des approches de résolution basées sur l'algorithme mimétique pour la résolution des problèmes d'affectation de fréquences modélisés en termes de jeux.

A ce moment, cinq (5) approches basées sur la théorie des jeux et l'algorithme mimétique pour résoudre différentes variantes du problème d'affectation de fréquence ainsi que la plateforme de programmation sont prêtes.

Un stage (de 7 mois) dans le cadre de bourse PNE a été effectué sous la supervision de Prof. Abdelmounaam Rezgui à New Mexico Tech. Ce dernier m'a permis d'enrichir mes connaissances à travers : une bibliographie très récente, les échanges avec l'encadreur et l'équipe de recherche à travers des présentations orales et des rencontres hebdomadaires, la participation aux événements scientifiques durant la durée du stage, accès à certaines ressources de l'institution d'accueil.

Comme résultat, une ébauche d'un premier article à publier très bientôt est en cours de développement (figure ci-dessous).

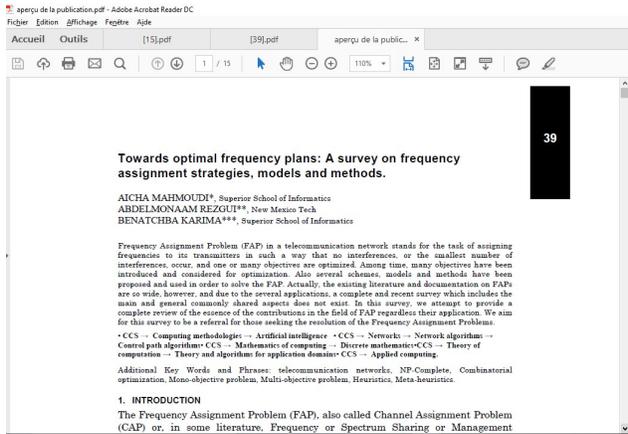


Fig. 1. Aperçu de l'article à soumettre

Cet article vient suite à la constatation que la documentation sur le PAF ne cesse d'accroître sans qu'il y ait un référentiel récent regroupant les travaux récents dans le domaine. D'autres articles de contribution suivront.

IV. - PERSPECTIVES

Ayant rencontré des problèmes de santé depuis le mois d'Avril 2017, mes travaux de thèse ont connu une interruption. Par conséquent, mon échéancier et objectif de soutenance de thèse a été retardé considérablement.

Mes objectifs actuels sont dans l'ordre :

- la finalisation et la soumission de l'article cité ci-dessus (au plus tard le 30 Septembre 2017).
- la validation des approches proposées à travers la publication d'articles dans des revues et/ou des conférences internationales (au plus tard Janvier 2018).
- la rédaction des chapitres de la thèse en attendant les notifications des articles soumis pour publication.

Sécurité et économie d'énergie dans l'Internet of Everything

Hamed Hellaoui

Directeurs de thèse : Abdelmadjid Bouabdallah, Mouloud Koudil

Équipe : CoDesign

Abstract—This report presents the progress state of my PhD thesis entitled 'Security and energy-efficiency in the Internet of Everything'. It highlights the issue tackled by the thesis and shows the work done till now. It also discusses the future direction of the thesis.

I. INTRODUCTION

The Internet of Things (IoT) emerges as a paradigm in which the network is extended to physical objects such as sensors, actuators, RFID tags, etc. It is considered as one of the most technologies affecting our daily lives that enabled many applications as smart cities, smart grids, health-care monitoring, etc. However, a widespread of this concept depends on its robustness and its security.

Many IoT applications are very sensitive. For instance, data measured by sensor devices in a health-care application may be related to human physiological signs which are very private. This highlights the requirement for applying security measures. In the other hand, security primitives are commonly known to be energy consuming. The IoT can use battery powered devices that are expected to operate for a long period. Security solutions must therefore be adapted to energy constraints of nodes in order to prolong their lifetime.

The issue of security and energy efficiency has been tackled in several networks such as the Wireless Sensor Networks (WSNs). However, the IoT comes with new characteristics that have to be taken into consideration. This thesis aims at contributing in the issue of security and energy-efficiency in the Internet of Things.

II. THESIS PROGRESS

Two goals have been established for the thesis.

- 1) A study on existing works dealing with the issue of security and energy efficiency needs to be performed.
- 2) Based on that study, a solution considering the IoT characteristics has to be proposed.

The next two subsections provide the conducted work to achieve the two goals.

A. Stat-of-the-art

With the emergence of Low-power and Lossy Networks (LLNs), several energy-efficient security protocols have been proposed. As energy is critical in such networks, existing surveys focus also in reviewing such solutions. However, no work has focused on what makes a security solution energy-efficient. Such a study, that interests in the mechanisms which

can be applied to make a security solution energy-efficient, would go further than just reviewing existing protocols. It could also help security protocol designers to develop energy-efficient security solutions by applying the mechanisms depending on their context. It is with this aim in mind that a survey on energy-efficient mechanisms in security of the Internet of Things is proposed. A taxonomy and classifications of the possible mechanisms are introduced. This work has been published in an international journal ranked A [1]. As shown in Figure 1, it tackles the following:

On-line/off-line security: It consists in absorbing a part of the security service by pre-computing and storing it. This part has to be calculable in advance and it will be used later.

Outsource security: It is based on delegating some parts of the security operations to more powerful nodes (resource-efficient devices). These parts are supposed to be more computationally intensive so to reduce the overhead on the constrained node.

Adaptive security: Static security maintains the same level. As being unchanged it must always consider the highest level, which is generally energy consuming. Adaptive security is based on adjusting the level of security depending the context. This would reduce the energy consumption compared to considering always the highest level.

Implementing using low-power security protocols: The emergence of LLNs led to the development of lightweight security protocols. This includes symmetric encryption, asymmetric encryption, as well as physical layer security protocols which are not based in encryption. Implementing using such protocols is very energy-efficient compared to old ones.

Size compression: Size compression techniques aim at reducing the consumption by decreasing the data size. As security services deal with data, the execution time is proportional to the data size. The more data are big, the more time is taken in the execution.

B. Solution

A special interest has been made to adaptive security. Indeed, a great energy will be wasted if the security level is high than what it is required. An efficient way to reduce energy consumption would start by not wasting it. The principle of adaptive security is based on adjusting the security level depending the context, rather than applying it systematically in the same manner. Therefore it would be necessary to identify the required level for each situation so to adapt security consequently.

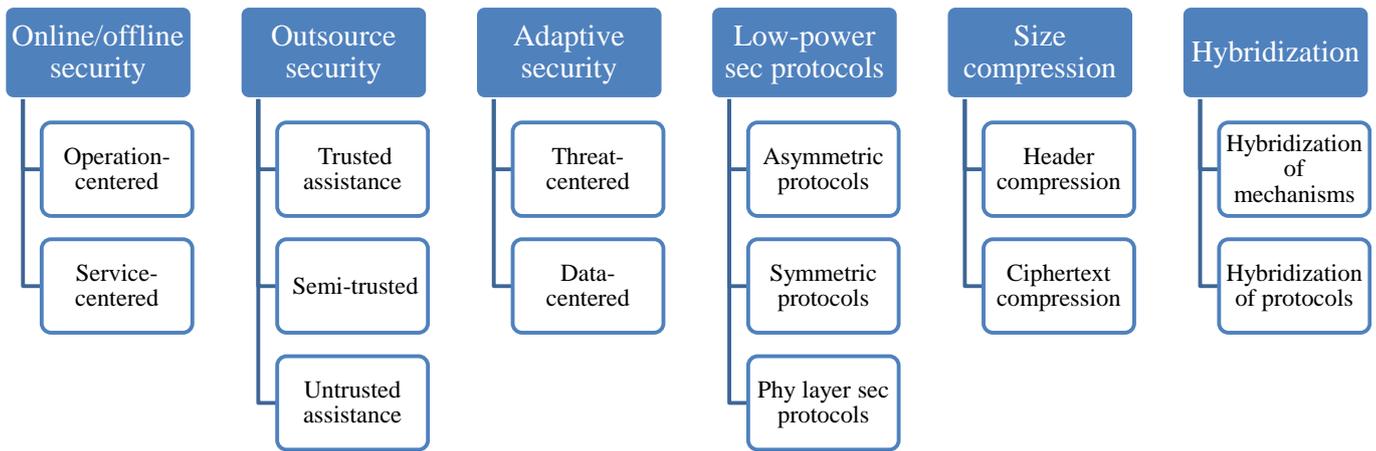


Fig. 1. The proposed taxonomy of energy-saving mechanisms in security protocols

A security model where nodes perform data authentication has been considered. This prevents a non-desired node from injecting packets in the network. However, applying this security service every time a packet is received would consume the node's energy. This consumption is considered a waste in cases where there is no need to authenticate the data.

A solution based on trust management is proposed to determine the need for the authentication service. This concept, trust management, has been widely used to deal with selfish behaviors and internal attacks in many networks. The provided solution allows reducing the application of the authentication service when nodes trust each other. This work has been published in an international conference ranked A [2].

III. FUTURE WORK

As the first goal of the thesis has been achieved, our entire focus is on completing the proposed solution. The following reflection is therefore made. The proposed solution allows performing adaptation during the use of the security service. However, adaptation could be also considered before the use of the security service (since the key establishment phase). Indeed, each cryptographic key, used in a security service, is associated with a lifespan. The latter represents the time required to recover the key using the most efficient method. For instance, in Elliptic Curve Cryptography (ECC), Pollard's rho algorithm can be used to recover the private key from the public one. The key size defines at the same time the level of security and the use duration of the service.

In the other hand, the IoT is a very dynamic environment. Its dynamic can be related to the node's stay in the system. Indeed, some devices may join the system for a fixed duration (such as in some group communication applications). Other devices that are battery-powered will leave the system once their battery is exhausted. Others may stay for a long duration. Therefore, it would be very interesting if the key generation phase is adapted so as to meet the node's stay.

As shown in Figure 2, different object types can be considered in the IoT with different stay duration. The key generation entity, responsible for generating the cryptographic keys, could take this parameter into consideration and adapt the key size for each node consequently. The key size has to

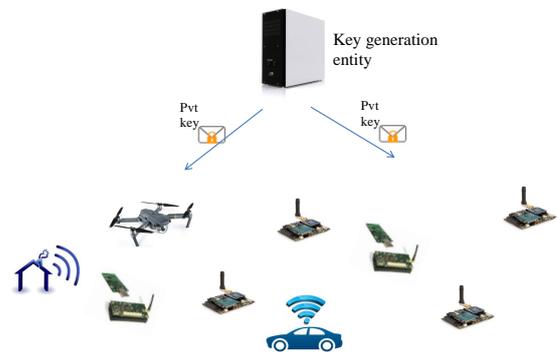


Fig. 2. Different object types can be considered in the IoT with different stay duration

be identified so to cover the node's stay and not to exceed it too much. Different strategies could be considered. This would allow adapting security at the key establishment phase (depending on the node's stay) and the use phase (depending on the node's trust).

IV. CONCLUSION

As energy is very critical for the Internet of Things, security services have to cope with this characteristic in order to prolong the network lifetime. The issue of security and energy-efficiency is tackled in this thesis. A survey on mechanisms that can be used to reduce energy consumption in security solutions is proposed. A solution based on adaptive security principle is also introduced. Completing the proposed solution is mandatory to finalize this thesis.

REFERENCES

- [1] H. Hellaoui, M. Koudil, and A. Bouabdallah, *Energy-efficient Mechanisms in Security of the Internet of Things: A survey*, Elsevier Computer Networks, 2017.
- [2] H. Hellaoui, A. Bouabdallah and M. Koudil, *TAS-IoT: Trust-based Adaptive Security for the Internet of Things*, IEEE LCN, 2016.

Résolution du problème d'affectation de fréquences multi objectifs dynamique en utilisant plusieurs approches biomimétiques

Asma DAOUDI

Directeurs de thèse : Professeur Karima BENATCHBA

Équipe : Optimisation

Abstract—Les objectifs de ce document sont d'une part la présentation de l'état d'avancement de l'implémentation des trois approches biomimétique appliqué sur le problème d'affectation de fréquence, tel que la DBO, BBO et NSGAI, et d'autre part l'analyse des résultats obtenus et attendu de notre implémentation. Dans un premier lieu nous allons présenter notre problématique de recherche dans la section 2, puis brièvement les travaux réalisés dans la section 3 suivi par une liste de tâches restantes pour la finalisation de notre thèse sous forme de perspectives en section 4.

I. INTRODUCTION

Le problème d'affectation de fréquences (FAP pour Frequency Assignment Problem) est l'un des problèmes d'optimisation combinatoires les plus émergents dans le domaine de la télécommunication en général et de la téléphonie mobile en particulier. Il consiste à affecter des fréquences aux stations d'un réseau de communication sans fil tout en respectant certaines contraintes.

Généralement, ces contraintes sont contradictoires et on est sensé les satisfaire simultanément, c'est pour quoi on a fait appel à l'optimisation multi-objectif qui s'intéresse à la résolution de ce type de problème.

Comme le FAP est un problème NP-difficile, il n'existe pas d'algorithme de résolution exacte, c'est pour cette raison que les méta-heuristiques jouent un rôle très important pour la résolution de ce problème. Dans la section suivante nous allons vous expliquer la difficulté de notre problème et comment remédié à ça dans le domaine d'optimisation combinatoire ce qui résume notre problématique.

II. PROBLÉMATIQUE

D'une manière concise, un bon plan fréquentiel doit permettre d'établir le maximum de communications en utilisant le minimum de fréquences et en minimisant les interférences ; ceci définit tout un domaine de recherche, c'est celui du problème d'affectation de fréquences PAF (frequency assignment problem, FAP).

D'une autre part, l'optimisation combinatoire occupe une place de plus en plus grande dans le monde de la recherche scientifique, vu son importance dans la modélisation, l'évaluation de la complexité, et la résolution de nombreux problèmes issus des différents domaines de la vie pratique, et qui sont supposés intraitables en raison de leur temps de résolution qui croit exponentiellement avec leur taille si on applique des méthodes visant à déterminer leur solution la plus optimale. Les travaux dans le domaine de la recherche opérationnelle

sur l'optimisation combinatoire, étant en pleine expansion, ça a donné naissance aux méta-heuristiques, qui sont des méthodes de résolution approchées, générales mais adaptables à de nombreux problèmes difficiles. Ceci a permis de donner des résultats satisfaisants quant à la résolution de ce dernier type de problèmes.

III. TRAVAUX RÉALISÉS

Après avoir fait une étude bibliographique, les années précédentes, sur les différents volets de recherche nécessaire à la résolution du problème d'affectation de fréquences multi objectives MO-PAF, nous avons proposé une modélisation du problème avec une multi T-coloration par liste. Les résultats obtenus sont très satisfaisants. Nous avons implémenté également une deuxième approche BBO (Biogeography Based Optimisation) sur notre plateforme pour un but de comparaison selon différentes métriques de performances. Les résultats d'exécutions sont collectés et validés. Pour valoriser les résultats obtenus, nous avons implémenté également une troisième approche spécifique au domaine multi-objectives à savoir NSGAI (Non-dominated Sorting Genetic Algorithm-II). La collecte des résultats a été faite et les différentes métriques de performance ont été calculées. Une étude non paramétrique basée les métriques de performance précédemment calculées pour chaque approche a fait l'objet d'un papier journal en cours de soumission. Enfin, un plan de rédaction de la thèse a été discuté.

IV. PERSPECTIVES

Après soumission de la version journal de notre travail de thèse, nous avons comme perspective de présenter une partie des résultats des expérimentations réalisées dans une conférence spécialisée dans le domaine d'intelligence artificiel. Ainsi, Nous envisageons d'explorer d'autres domaines d'application des problèmes multi-objectives et adapter notre approche à l'un des ces domaines. à savoir, la coloration des graphes et l'Internet of Things (IoT).

V. CONCLUSION

Dans le présent document nous avons essayé de synthétiser l'état d'avancement de notre thèse. Dans un premier lieu nous avons positionné notre problématique dans la deuxième section, suivie par un résumé des travaux déjà réalisé, depuis l'état de l'art jusqu'à l'implémentation dans la troisième section. Dans la dernière et la quatrième section nous avons listé l'ensemble des tâches envisagés pour l'année en cours comme perspectives.

REFERENCES

- [1] <http://jmetal.sourceforge.net/>

Modèles et Algorithmes Skyline pour la Sélection des Services Web Basés sur l’Historique d’Invocation

Abdelaziz OUADAH

Directeurs de thèse : Allel HADJALI, Fahima NADER

Équipe : LCMS, ESI, Alger.

Résumé— Ce papier décrit l’état d’avancement dans les travaux de thèse ; il présente la problématique traitée, l’approche proposée, les papiers publiés ainsi que les perspectives projetées.

I. INTRODUCTION

Les services web sont des composants logiciels qui peuvent être publiés et invoqués sur le web utilisant un ensemble de standards comme SOAP, WSDL et UDDI. Chaque service est conçu dans le but de réaliser une tâche bien précise. Un service est décrit par son identificateur sur le Web et ses paramètres d’entrée et de sortie qui représentent ses informations fonctionnelles, cependant, il existe d’autres informations non fonctionnelles qui peuvent caractériser le service Web comme le temps de réponse, la disponibilité, le prix du service, la fiabilité, ...etc., ces informations sont appelées, qualité de service (qos).

II. PROBLEMATIQUE

Pour sélectionner un service parmi plusieurs services candidats, qui offrent la même fonctionnalité mais avec des qualités de services divers, l’utilisateur ne peut pas tester tous les services candidats. Par contre, il peut se baser sur l’historique des précédentes invocations effectuées par d’autres utilisateurs dans des contextes similaires [1]. Ces dernières années, la notion de Skyline [2] est émergée comme étant un nouveau et populaire paradigme pour trouver les services Web les plus pertinents [3], [4]. Il s’agit d’une direction prometteuse qui permet de réduire l’espace de décision de l’utilisateur en lui proposant uniquement les services les plus intéressants, et par conséquent, simplifier le processus de sélection. Cependant, Skyline présente quelques inconvénients : i) elle retourne des résultats incomparables, ou l’utilisateur trouve des difficultés pour exprimer ses préférences et choisir les services pertinents avec un meilleur compromis entre les critères, ii) Skyline accorde la même importance à tous les critères de sélection, alors que cet importance diffère d’un contexte à un autre et d’un domaine à un autre, par exemple dans le domaine de les services e-paiement le critère *sécurité* est plus important que *temps de réponse*, iii) le résultat Skyline ne peut pas être contrôlé, en fait, Skyline peut retourner un nombre très élevé de services, ce qui rends le processus de sélection très lourd et pas pratique.

III. TRAVAUX REALISES

A. Proposition 01 : Combiner Skyline avec les méthodes MCDM pour améliorer la sélection des services

Nous avons proposé une approche pour la sélection des services web appuyant sur l’historique d’invocation effectué par d’autres utilisateurs [5][6], cette approche est basée sur la

combinaison de plusieurs méthodes d’aide à la décision multicritère à savoir : Skyline, AHP et PROMETHEE. Dans la première étape de notre approche, nous utilisons la méthode Skyline pour réduire l’espace de recherche de l’utilisateur. Ensuite, nous calculons les poids des critères de Qos appuyant sur la méthode AHP qui offre la possibilité aux utilisateurs d’exprimer leurs préférences sous forme de comparaison binaire entre chaque deux critères d’une manière simple et intuitive, minimisant l’effort de l’utilisateur et garantissant que les poids assignés aux critères sont cohérents. Dans la dernière étape, nous proposons d’utiliser la méthode Prométhée pour faire un rangement des services web Skyline, proposant à l’utilisateur les services qui ont un bon compromis entre les critères de qos, exploitant les relations de surclassement, les différentes fonctions d’expression des préférences sur les critères telles que les indicateurs de préférence, les indicateurs d’indifférence, le sens des préférence et les poids des critères calculés par le biais d’une autre méthode tel que AHP dans notre cas.

Durant cette année nous avons travaillé pour enrichir notre approche proposée dans [5] et [6] afin de traiter l’incertitude dans jugements de l’utilisateur, nous avons utilisé la méthode objective basée sur les données « Entropie » et la méthode subjective basée sur le jugement de l’utilisateur « FuzzyAHP » pour calculer les poids des critères de Qos. Des expérimentations ont été réalisées avec des comparaisons avec des approches existantes [8][9].

Ce travail a été envoyé pour publication dans la revue : « ACM Transactions on the Web (TWEB) », il a été rejeté. Des modifications ont été apportées sur l’article, il a été renvoyé à une autre revue Springer « Journal of Ambient Intelligence & Humanized Computing »(AIHC, indexé par SCOPUS et Thomson Reuters), des révisions nous ont été demandées, elles sont en cours de prise en considération.

B. Proposition 02 : sélection des K-représentative skyline services.

La deuxième proposition [7](voir figure 1)consiste à réduire l’espace de décision de l’utilisateur en lui proposant uniquement les services Skyline qui représentent un sommaire de tout le résultat Skyline [10][11], l’approche consiste à segmenter le résultat Skyline en k segments utilisant K-means, et de prendre de chaque segment un service qui représente le segment, puis, les k-représentative Skyline services sont rangés selon les préférences de l’utilisateur (poids des critères, indice de préférence et d’indifférence). Les poids des critères de Qos sont calculés utilisant une version améliorée de la méthode Fuzzy AHP, groupant les critères qui ont une importance similaire exprimé par l’utilisateur.

Une version améliorée du papier [7] est en cours de rédaction.

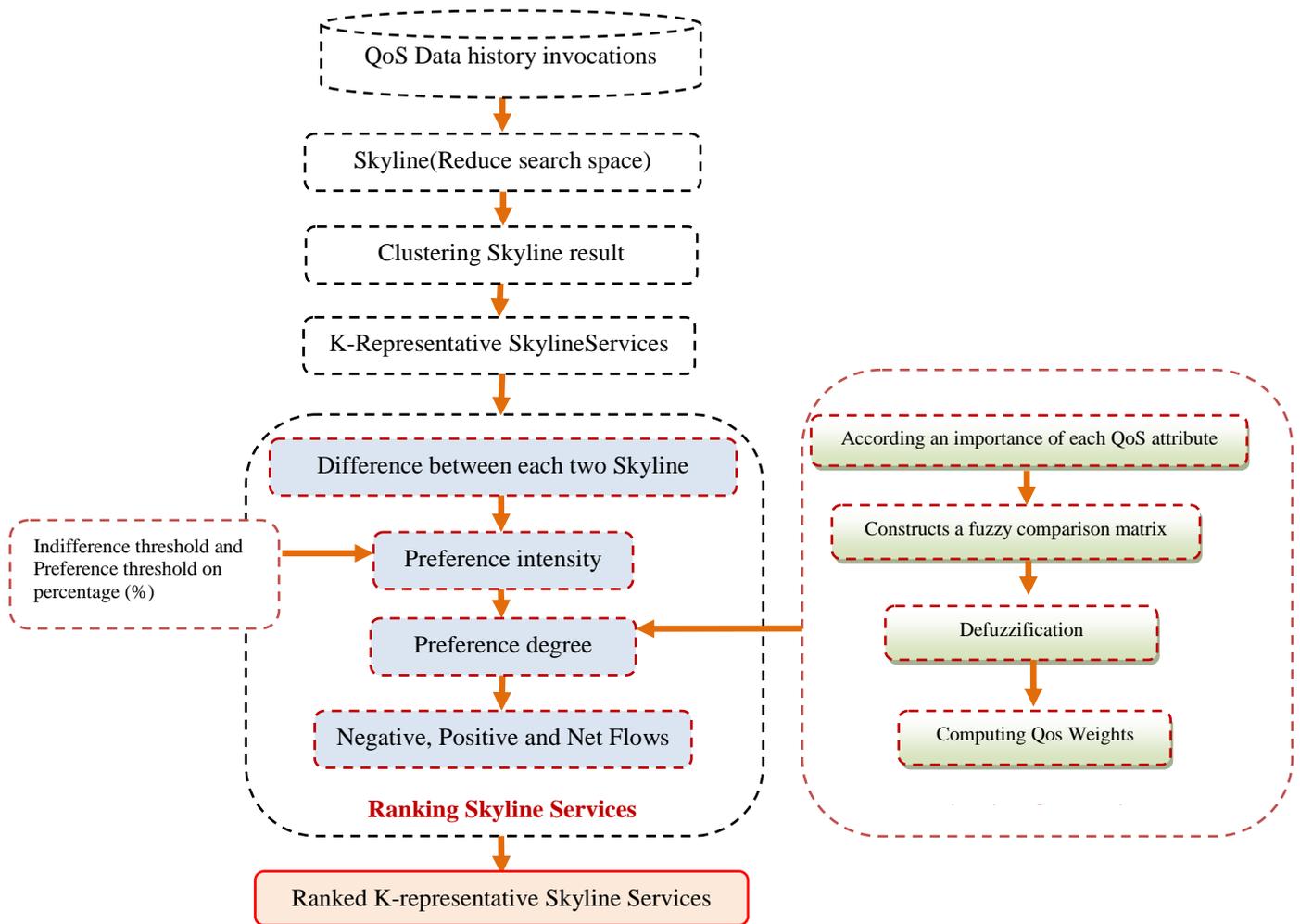


Figure 1. Approche pour la sélection des K-représentative Skyline Web Services basée sur l’historique d’invocation.

IV. CONCLUSION

Durant cette année nous avons travaillé sur l’enrichissement des papiers publiés dans des conférences concernant la première proposition afin de publier un article de revue. L’article de revue est en cours de révision par le Journal of Ambient Intelligence & Humanized Computing (AIHC). Nous projetons l’année prochaine d’effectuer des expérimentations approfondies sur l’approche (figure 1) [7] et de réaliser une étude comparative approfondie avec les approches existantes.

V. REFERENCES

- [1] Shao, J. Zhang, There. Wei, J. Zhao, B. Xie, and H. Mei. “Personalized QoS prediction for web services via collaborative filtering”. In ICWS, pages 439-446, on 2007. DOI=[10.1109/ICWS.2007.140](https://doi.org/10.1109/ICWS.2007.140)
- [2] S. Borzsonyi, Orzs Onyi, D. Kossmann, and K. Stock. "The Skyline operator". IN ICDE. On 2001.
- [3] K. Benouaret, Benslimane, Djamal, and Hadjali, Allel. “Selecting Skyline Web Services for Multiple Users Preferences”. In: Web Services (ICWS), 2012 IEEE 19th International Conference one. IEEE.p. 635-636. On 2012.DOI= [10.1109/ICWS.2012.108](https://doi.org/10.1109/ICWS.2012.108)
- [4] Q. Yu and Athman Bouguettaya, “Efficient Foundations for Web Selection service”. Springer, 2010.
- [5] A. Ouadah, K. Benouaret, A. HADJALI, F. Nader , “Combining Skyline and Multi-Criteria Decision Methods to Enhance Web Services Selection”, 12th International Symposium on Programming and Systems (ISPS 2015), Algiers, Algeria, April 28-30 2015.DOI=[10.1109/ISPS.2015.7244975](https://doi.org/10.1109/ISPS.2015.7244975)
- [6] A. Ouadah, K. Benouaret, A. HADJALI, F. Nader, “SkyAP-S3: A Hybrid Approach for Efficient Skyline Services Selection”, Proc. of the 8th IEEE International Conference on Service Oriented Computing & Applications (SOCA 2015), Rome, Italy, 19-21 October 2015.DOI=[10.1109/SOCA.2015.22](https://doi.org/10.1109/SOCA.2015.22)
- [7] A. Ouadah, A. Hadjali, F. Nader. “An Efficient Framework for Selecting K-representative Skyline Services”, 3rd ACM International Conference of Computing for Engineering and Sciences (ICCES'2017), Istanbul, Turkie, 22-24 July 2017 (Accepted).
- [8] G. Dai, Qingsheng Zhu. “Using Skyline and Dominance Relationship for Web Ranking services”. Journal of Computational Information Systems 9: 10 (3977-3984). On 2013.
- [9] Vlachou Akrivi, and Michalis Vazirgiannis. "Ranking the sky: Discovering the importance of skyline points through subspace dominance relationships." Data & Knowledge Engineering 69.9 (2010): 943-964.DOI=[10.1016/j.datak.2010.03.008](https://doi.org/10.1016/j.datak.2010.03.008)
- [10] Lin, X., Yuan, Y., Zhang, Q., & Zhang, Y. (2007, April). Selecting stars: The k most representative skyline operator. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 86-95). IEEE. DOI= <https://doi.org/10.1109/ICDE.2007.367854>
- [11] Tao, Y., Ding, L., Lin, X., & Pei, J. (2009, March). Distance-based representative skyline. In Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on (pp. 892-903). IEEE. DOI= <https://doi.org/10.1109/ICDE.2009.84>

Analyse des réseaux sociaux pour la détection d'harcèlement

Nihed Bader

Directeurs de thèse : Faycal Azouaou

Équipe : (LMCS)

Résumé— Internet se révèle plus que jamais un outil indispensable d'échange d'informations. Il a connu encore une plus vaste portée grâce à l'émergence des réseaux sociaux. Malgré leurs rôles importants, ces réseaux peuvent devenir un danger où certaines informations pourraient être mal interprétées et utilisées à mauvais escient, d'où le « cyber harcèlement ». L'objectif de notre travail est d'utiliser différentes méthodes automatiques et statistiques permettant d'analyser les données provenant des réseaux sociaux afin de pouvoir détecter les contenus (ainsi que leurs auteurs) qui sont les plus préjudiciables. Cet article représente une description des travaux de recherches réalisés durant la 1ère année de thèse, ainsi que les perspectives d'avenir.

I. INTRODUCTION

Outre l'objectif principal du web, qui est d'assurer le libre accès à l'information, la nouvelle génération de réseaux sociaux a introduit aux utilisateurs de nouvelles libertés dans leurs relations avec le Web, en leur permettant d'interagir avec les autres utilisateurs, de générer du contenu et de donner leurs opinions sur divers sujets. Ainsi, les réseaux sociaux sont devenus une cible privilégiée pour les différentes actions malveillantes comme le cyber harcèlement.

L'objectif de notre travail est de proposer une méthode outillée qui permet de détecter le harcèlement dans les réseaux sociaux. Pour cela, nous comptons exploiter les techniques de la fouille de données appliquées aux médias sociaux (social mining), analyse et traitement du texte pour pouvoir détecter et prévenir le harcèlement dans les réseaux sociaux.

II. PROBLEMATIQUE

Un des problèmes les plus inquiétants des réseaux sociaux est le cyber harcèlement : le fait d'harcéler une personne ou de tenir à son endroit des propos menaçants, haineux, injurieux ou dégradants, qu'ils soient illustrés ou écrits via les réseaux sociaux. Il peut prendre plusieurs formes telles que : les intimidations, insultes, moqueries ou menaces en ligne, la propagation de rumeurs, la création d'un sujet de discussion, d'un groupe ou d'une page sur un réseau social à l'encontre d'une personne ou encore la publication d'une photo ou d'une vidéo de la victime en mauvaise posture.

Selon l'étude de Pew Reserarch [01] : 73% des internautes ont observé un harcèlement et 40% l'ont vécu, parmi eux, 66% l'ont subi sur les réseaux sociaux. Le cyber harcèlement engendre plusieurs conséquences allant de la perte de l'estime de soi à la dépression et même, dans certains cas, au suicide.

III. TRAVAUX REALISÉS

Face aux sérieuses conséquences qu'engendre le cyber harcèlement, il importe de concevoir un système automatique apte à la recherche et l'analyse des données exprimées sur les réseaux sociaux afin de pouvoir détecter et diminuer ce danger. Dans ce contexte, un projet intitulé « Analyse des réseaux sociaux pour la détection d'harcèlement », vise à dresser un état de l'art sur les techniques d'analyse de l'information sociale et de proposer un outil aidant à la détection des individus à comportement suspect d'harcèlement.

La première étape dans cette étude consiste à effectuer une recherche des travaux antécédents dans la détection d'harcèlement afin de pouvoir combler les lacunes de ces derniers. Nous avons étudié plusieurs travaux publiés dans ce domaine [2-19], nous citons :

- ❖ L'un des premiers travaux à aborder la détection du langage abusif était de Yin et al. [02], utilisant des modèles d'expressions développés manuellement et tenant compte des phrases précédentes d'abus.
- ❖ Le travail de Sood et al. [03], utilisant la notion de liste noire améliorée par l'ajout de mots non normalisés par un groupe de travailleurs de Amazon Mechanical Turk qui avaient étiqueté 6500 commentaires sur Internet.
- ❖ Le travail de Xu et al. [04], appliquant des modèles du traitement du langage naturel (BOW, Latent Semantic Analysis (LSA) et Latent Dirichlet Allocation (LDA)) pour la détection d'harcèlement dans les réseaux sociaux.
- ❖ Le travail de Nahar et al. [05], combinant deux techniques du traitement du langage naturel : Probabilistic Latent Semantic Analysis (PLSA) et Latent Dirichlet Allocation (LDA), pour la détection du contenu abusif.
- ❖ Le travail de Djuric et al. [06], qui se base sur une approche paragraph2vec pour classer les commentaires des utilisateurs comme abusive ou non. Il implémente la méthode de sac à mots (BOW).
- ❖ Le travail d'Intel's #HackHarassment groupe [07] qui propose une variété de modèles types testés sur une base de données de commentaires et de publications que leur équipe avait étiqueté.
- ❖ Le travail de Pater et al. [08], proposant l'étude des politiques de sécurités des différentes plateformes de réseaux sociaux afin de limiter le harcèlement.
- ❖ Le travail de Guberman et al. [09] proposant la détection d'harcèlement dans les tweets.
- ❖ Le travail de Kennedy et al. [10], qui constitue une base de données multiplateforme et un modèle d'apprentissage par machine construit sur l'ensemble de ces données.

Encore et récemment, plusieurs compagnies d'internet ont concentré leurs efforts sur la lutte contre les diverses formes d'harcèlement en ligne, parmi lesquelles nous citons :

- ❖ Le travail des chercheurs de Yahoo [11], qui avaient développé des modèles d'apprentissage par machine pour détecter un langage abusif.
- ❖ Une étude de l'équipe Google Jigsaw en partenariat avec la fondation Wikimedia [12] pour développer des solutions servant de réduire les commentaires toxiques dans Wikimedia.

IV. SYNTHÈSE

Tous ces travaux visent à détecter le harcèlement dans les réseaux sociaux en se basant uniquement sur le contenu et néglige un facteur très important qui est « le contexte » : un message utilisant un langage abusif et échangé entre deux amis ne peut être considéré comme harcèlement. Encore, un message envoyé une seule fois, ou qui utilise des propos pas trop abusifs ne peut pas être considéré comme harcèlement.

Dans le cadre de notre recherche, nous nous s'intéressons à trois facteurs importants dans la détection d'harcèlement : le **contexte**, le **degré d'harcèlement** et sa **fréquence**. En s'appuyant sur ces trois facteurs, la décision qu'un contenu est un harcèlement ou non peut être prise.

Dans le cadre de notre étude, nous nous s'intéressons au domaine du 'Online Social Network Mining' [20,23] : des techniques de la fouille de données appliquées aux médias sociaux permettant l'analyse et le traitement du texte. Cette analyse des interactions entre les utilisateurs des réseaux donne la possibilité de la prédiction en temps réel de l'évolution de l'information et la détection des contenus (ainsi que leurs auteurs) qui sont les plus préjudiciables, ce qui permet, par conséquent, de pouvoir détecter et prévenir le harcèlement dans les réseaux sociaux. Ce domaine fera l'objet de notre prochaine phase de recherche et d'étude.

V. PERSPECTIVES

Nous planifions pour l'année prochaine de finaliser en notre phase de recherche par une étude des différentes techniques de fouille de données dans les réseaux sociaux. Ensuite nous comptons finaliser la rédaction de l'état de l'art portant sur tous les domaines d'étude et proposer par la suite une approche de détection qui fera le sujet d'une communication. A long termes les travaux qui nous restent sont :

- Evaluation de l'approche.
- Rédaction et soumission de la publication.
- Rédaction de la thèse.

VI. CONCLUSION

À travers ce papier, nous avons présenté l'état d'avancement de nos travaux de thèse pour la première année de recherche. Après un rappel du sujet et de la problématique traitée, nous avons présenté brièvement les travaux réalisés ainsi que les perspectives fixées pour les années à suivre.

References

- [1] Duggan, M., Rainie, L., Smith, A., Funk, C., Lenhart, A., and Madden, M. *Online Harassment*. Pew Research Center, 2014
- [2] Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A., & Edwards, L. (2009). Detection of harassment on web 2.0. *Proceedings of the Content Analysis in the WEB*, 2, 1-7.
- [3] Sood, S. O., Antin, J., & Churchill, E. F. (2012, March). Using Crowdsourcing to Improve Profanity Detection. In *AAAI Spring Symposium: Wisdom of the Crowd*(Vol. 12, p. 06).
- [4] Xu, J. M., Jun, K. S., Zhu, X., & Bellmore, A. (2012, June). Learning from bullying traces in social media. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 656-666). Association for Computational Linguistics.
- [5] Nahar, V., Li, X., & Pang, C. (2013). An effective approach for cyberbullying detection. *Communications in Information Science and Management Engineering*, 3(5), 238.
- [6] Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., & Bhamidipati, N. (2015, May). Hate speech detection with comment embeddings. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 29-30). ACM.
- [7] Bastidas, A., Dixon, E., Loo, C., & Ryan, J. (2016). Harassment detection: a benchmark on the# HackHarassment dataset. *arXiv preprint arXiv:1609.02809*.
- [8] Pater, J. A., Kim, M. K., Mynatt, E. D., & Fiesler, C. (2016, November). Characterizations of Online Harassment: Comparing Policies Across Social Media Platforms. In *GROUP* (pp. 369-374).
- [9] Guberman, J., & Hemphill, L. (2017, January). Challenges in modifying existing scales for detecting harassment in individual tweets. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.
- [10] Kennedy, G., McCollough, A., Dixon, E., Bastidas, A., Ryan, J., Loo, C., & Sahay, S. (2017). Technology Solutions to Combat Online Harassment. In *Proceedings of the First Workshop on Abusive Language Online* (pp. 73-77).
- [11] Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016, April). Abusive language detection in online user content. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 145-153). International World Wide Web Conferences Steering Committee.
- [12] E. Wulczyn, D. Taraborelli, N. Thain, and L Dixon. (2017). Algorithms and insults: Scaling up our understanding of harassment on wikipedia.
- [13] Chen, H., McKeever, S., & Delany, S. J. (2017). Harnessing the Power of Text Mining for the Detection of Abusive Content in Social Media. In *Advances in Computational Intelligence Systems* (pp. 187-205). Springer International Publishing.
- [14] Zhao, R., Zhou, A., & Mao, K. (2016, January). Automatic detection of cyberbullying on social networks based on bullying features. In *Proceedings of the 17th international conference on distributed computing and networking* (p. 43). ACM.
- [15] Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving Google's Perspective API Built for Detecting Toxic Comments. *arXiv preprint arXiv:1702.08138*.
- [16] Dadvar, M., Trieschnigg, D., & de Jong, F. (2014, May). Experts and machines against bullies: A hybrid approach to detect cyberbullies. In *Canadian Conference on Artificial Intelligence* (pp. 275-281). Springer, Cham.
- [17] Mangaonkar, A., Hayrapetian, A., & Raje, R. (2015, May). Collaborative detection of cyberbullying behavior in Twitter data. In *Electro/Information Technology (EIT), 2015 IEEE International Conference on* (pp. 611-616). IEEE.
- [18] Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, 7(2), 223-242.
- [19] Hosseinmardi, H., Mattson, S. A., Rafiq, R. I., Han, R., Lv, Q., & Mishra, S. (2015). Detection of cyberbullying incidents on the instagram social network. *arXiv preprint arXiv:1503.03909*.
- [20] Ristoski, P., & Paulheim, H. (2016). Semantic Web in data mining and knowledge discovery: A comprehensive survey. *Web semantics: science, services and agents on the World Wide Web*, 36, 1-22.
- [21] KS, Deepashri; KAMATH, Ashwini.(2017). Survey on Techniques of Data Mining and its Applications.
- [22] Kurka, D. B., Godoy, A., & Von Zuben, F. J. (2015). Online social network analysis: A survey of research applications in computer science. *arXiv preprint arXiv:1504.05655*.
- [23] Injadat, M., Salo, F., & Nassif, A. B. (2016). Data mining techniques in social media: A survey. *Neurocomputing*, 214, 654-670.

Ordonnancement conjoint de la production et de la maintenance avec contrainte d'affectation des ressources humaines dans les ateliers de production

Meriem TOUAT

Directeurs de thèse : Fatima Benbouzid-Si Tayeb, Belaid Benhamou

Équipe : (Optimisation)

Résumé— L'objectif de la thèse est de proposer des approches originales d'ordonnancement conjoint de production et de maintenance qui prennent en compte les contraintes liées à la compétence et à la planification du travail des ressources humaines de la maintenance. Plusieurs algorithmes génétiques ont été proposés pour résoudre le problème ces années passées. Le présent papier présente l'état d'avancement de la troisième année de notre thèse, ensuite les perspectives envisagées pour la 4^{ème} année.

I. INTRODUCTION

La production et la maintenance sont deux fonctions pivots de l'entreprise qui agissent sur les mêmes ressources. Cependant leurs objectifs sont antagonistes : d'un côté la production qui a pour objectif d'optimiser la triptyque (qualité, coût et prix) pour une meilleure rentabilité de l'outil de production ; de l'autre la maintenance qui vise à augmenter la disponibilité de l'outil de production en assurant une meilleure qualité d'entretien. Ceci nous amène à penser qu'il est préférable d'intégrer les activités de maintenance dans l'ordonnancement de la production et d'établir un ordonnancement conjoint de production et de maintenance.

Les activités de production et de maintenance font appel à des ressources matérielles et humaines pour leur achèvement. Les ressources humaines sont des opérateurs qui veillent au bon déroulement des tâches de production et de maintenance. Leur rôle peut varier d'un contrôle simple du déroulement des tâches à une intervention manuelle sur les machines. Ceci dépendra du type des machines, des produits et des tâches à réaliser.

Dans ce résumé étendu, nous rappelons la problématique abordée dans la thèse dans un premier temps. Ensuite, nous présentons les travaux réalisés au cours de la troisième année. Les perspectives envisagées pour notre quatrième année seront détaillées par la suite et nous clôturons par une conclusion.

II. PROBLEMATIQUE

L'ordonnancement conjoint de production et de maintenance a fait l'objet de plusieurs travaux intéressants. Cependant, les tâches de production et les activités de maintenance sont ordonnancées de la même façon malgré le fait qu'elles ne possèdent pas les mêmes contraintes. En effet, les activités de maintenance font intervenir des ressources humaines ayant différentes compétences, différentes disponibilités et différents coûts. Ces propriétés doivent être prises en compte pour générer un ordonnancement conjoint réaliste. Les travaux existants ne considèrent que la contrainte de compétence et ne se soucient pas des contraintes de planification du travail (disponibilité, équité et coût). par ailleurs, les tâches de production nécessitent la présence des

ressources humaines pour le bon déroulement du processus de production. Les travaux existants considèrent un rôle de contrôle pour les ressources humaines et essaient de résoudre le problème conjoint d'affectation des ressources humaines et d'ordonnancement des tâches de production en prenant en compte les contraintes de planification du travail (réglementation, équité et préférences).

L'objectif donc de cette thèse est de proposer des approches originales d'ordonnancement conjoint de production et de maintenance qui prennent en compte les contraintes liées à la compétence et à la planification du travail des ressources humaines de la maintenance. En effet, la compétence peut être considérée non seulement pour chercher l'opérateur le plus performant mais d'essayer par exemple de faire intervenir les débutants afin d'améliorer leur compétence tout en assurant un minimum de satisfaction en ce qui concerne la qualité et les délais de réalisation de la tâche en question (production ou maintenance). La contrainte des règles de planification du travail, quant à elles, assurent la proposition d'ordonnements réalistes et applicables du moment que la disponibilité et l'équité des ressources humaines sont prises en compte. La contrainte du coût d'intervention à son tour présente un déficit pour les entreprises du fait qu'elles cherchent à réaliser l'équilibre entre la minimisation des coûts et l'amélioration de la qualité du service.

III. TRAVAUX REALISES

A. La résolution de la version « avec incertitudes » de la problématique

Nous avons abordé la problématique selon une première vision. En effet, nous avons considéré le contexte suivant:

- Ordonnancer conjointement des tâches de production et de maintenance flexibles.
- Prendre en compte les contraintes de disponibilité et de compétence des ressources humaine de la maintenance.
- Tenir compte les incertitudes dans le contexte des ateliers de production, en particulier, les imprécisions des durées opératoires de la production et de la maintenance.
- Optimiser une fonction objectif linéaire commune à la production et à la maintenance

Pour résoudre la problématique nous avons proposé deux algorithmes génétiques (AG) flous selon deux approches de résolution dites séquentielle et intégrée où nous avons utilisé la logique floue pour modéliser les durées opératoires

imprécises. La première approche consiste à (1) générer une population d'ordonnancement conjoint sans tenir compte de la contrainte des ressources humaines. Ensuite (2) à affecter les ressources humaines aux ordonnancements selon leurs faisabilités (si la ressource humaine est disponible au moment de la planification de la maintenance, l'ordonnancement est faisable). La deuxième approche considère la contrainte des ressources humaines au moment de la génération de la population d'ordonnancement, d'où la nécessité d'un processus de réajustement après chaque opération de croisement ou de mutation de l'algorithme génétique.

Nous avons capitalisé ces travaux à travers deux publications:

- Une revue [1] "A hybridization of genetic algorithms and fuzzy logic for the single-machine scheduling with flexible maintenance problem under human resource constraints".
- Une conférence [2] "A Fuzzy Genetic Algorithm for Single Machine Scheduling and Flexible Maintenance Planning Integration under Human Resource Constraints". Où un autre algorithme génétique est proposé pour résoudre la même problématique.

B. La résolution de la version déterministe-exacte de la problématique

Cette tâche a été réalisée lors d'un stage de perfectionnement sous la direction de Mr Benhamou, elle peut être résumée ainsi:

- Formulation mathématique du problème en MILP (Mixed Integer Linear Programming) : spécification de toutes les contraintes du problème, optimisation du modèle (éliminer les redondances et les contraintes superflues, simplifier les équations afin d'optimiser les temps de calcul,...).
- Implémenter le modèle sous IBM ILOG CPLEX Optimization Studio : l'IDE étant le plus utilisé offre des panoplies d'outils facilitant l'implémentation des modèles ainsi qu'avoir des solutions en un temps de calcul très performant. Pour ce faire, l'apprentissage du langage OPL (Optimization Programming Language) dédié à l'implémentation des modèles mathématiques sous ILOG Cplex IBM est indispensable.
- Définir le protocole de tests en mettant en exergue les points sensibles et influents de la problématique.
- Résolution du modèle en utilisant le cluster du laboratoire et le solver ILOG Cplex d'IBM : en passant par l'apprentissage de Linux pour la gestion des flux dans un réseau, et par la résolution de tous les détails techniques de compatibilité entre : Linux/ILOG Cplex/ format des fichiers de données.

- Effectuer les expérimentations sur (5X10X4) cas dont certaines expérimentations ont duré plus de 6 heures (vu la complexité et la sensibilité du problème) et présenter les résultats sous forme interprétable.

C. La résolution de la version déterministe-approchée de la problématique

Le schéma de réalisation de cette tâche a été établi lors de mon stage de perfectionnement, il s'agit de proposer des heuristiques dédiées ainsi qu'adapter la méthode VNS (Variable Neighborhood Search) comme métaheuristique de résolution approchée.

IV. PERSPECTIVES

Pour l'année à venir, nous visons les perspectives suivantes:

A. Finaliser les tâches abordées dans l'année 3

En particulier, (1) les heuristiques et VNS, il sera question de finaliser l'implémentation et d'étudier les propriétés et comparer les résultats, (2) définir des bornes au problème.

B. Capitaliser les travaux réalisés

Nous visons de capitaliser les travaux réalisés à travers :

- Une publication dans une conférence de la méthode exacte et les heuristiques.
- Une version longue « publication dans une revue » de toutes les approches de résolution proposées : la VNS, les heuristiques, l'AG.

C. Rédiger le mémoire.

V. CONCLUSION

A travers ce résumé étendu nous avons présenté l'état d'avancement des trois années de la thèse. Nous avons proposé des algorithmes génétiques pour la résolution de la version avec incertitudes de la problématique, cette contribution a été capitalisée à travers deux publications de haut niveau. De plus, nous avons formulé mathématiquement la problématique en question afin de résoudre la version exacte du problème. D'autres horizons de recherche ont été définis notamment l'utilisation d'autres méthodes de résolution de la problématique.

References

- [1] M. Touat, S. Bouzidi-Hassini, F. Benbouzid-Si Tayeb, B. Benhamou "A hybridization of genetic algorithms and fuzzy logic for the single-machine scheduling with flexible maintenance problem under human resource constraints". Appl. Soft Comput. 59: 556-573 (2017).
- [2] M. Touat, F. Benbouzid-Si Tayeb, S. Bouzidi-Hassini, B. Benhamou "A Fuzzy Genetic Algorithm for Single Machine Scheduling and Flexible Maintenance Planning Integration under Human Resource Constraints", 29th International Conference on Tools with Artificial Intelligence (ICTAI), November 6-8, 2017, Boston, Mass, USA, Accepted but not yet published.

Implémentation de techniques avancées d'Interaction Homme Robot pour un robot de service

Souhila Kahlouche

Directeur de thèse : Belhocine Mahmoud

Centre de Développement des Technologies Avancées (CDTA)

Résumé— Dans l'état d'avancement de la première année, nous avons présenté l'histoire de l'émergence du domaine d'Interaction Homme Robot (IHR) à partir de l'Interaction Homme Machine (IHM). Nous avons par la suite défini les différents aspects à prendre en compte pour concevoir un système d'IHR. Dans ce rapport, nous allons situer notre sujet par rapport à ce vaste domaine de recherche et présenter la démarche à suivre pour atteindre nos objectifs.

I. INTRODUCTION

- L'Interaction Homme Robot est un domaine de recherche émergent, qui attire l'attention de beaucoup de chercheurs de différentes disciplines comme les roboticiens, les psychologues et les chercheurs dans l'intelligence artificielle.

La figure 1 représente une liste non exhaustive des axes de recherche dans le domaine de l'IHR :

- L'interaction verbale, qui permet le transfert d'information riche en contenu entre le robot et son interlocuteur, plusieurs travaux dans ce domaine ont été réalisés, comme le système d'interaction verbale pour les instructions culinaires [1], le système d'interaction pour les robots guides et d'exploration de ville [2], les systèmes d'interaction verbale pour des robots livreurs de snack [3] et les robots réceptionnistes [4].

- L'interaction non verbale, où le robot fusionne plusieurs modalités incluant : les gestes et mouvements du corps humain [5], les expressions faciales, et le regard [6], la gestion de l'espace et de proximité [7] et l'interaction tactile [8].

- La reconnaissance d'intention qui a pour objectif d'équiper le robot d'une compétence de haut niveau qui est l'inférence des intentions des humains et l'expression de ses propres intentions. Les travaux dans cet axe concernent l'analyse d'intention avec laquelle le robot raisonne sur l'espace de travail partagé avec son collaborateur humain [9], l'attention jointe est également une capacité très importante dans la manipulation partagée des objets, elle améliore les performances du robot lors de l'interaction [10].

- La navigation sociale; dans laquelle le robot doit traiter la navigation comme étant un problème social où il doit coordonner entre ses propres intentions (atteindre sa destination) et les intentions des humains, les travaux réalisés dans ce domaine concernent les robots approchant les humains [11], les robots navigant côte à côte avec les humains [12] et les robots poursuivant les humains [13].

- La perception de l'humain, qui a pour objectifs: la reconnaissance des humains et leurs poses [14], l'analyse de leurs actions [15], et la détection de leur engagement pour l'interaction [16]. Notre sujet s'insère dans cet axe, dans l'objectif de munir un robot de service de type robot guide, d'une interaction naturelle et intuitive. Il s'agit en fait, d'analyser et de reconnaître les activités des humains afin d'inférer leurs intentions et de les aider à atteindre leurs objectifs. Dans la figure 1, nous soulignons l'axe de recherche qui nous intéresse, notamment la perception de l'humain et particulièrement la détection de sa pose et la reconnaissance

de ses d'activités par le robot, et nous présentons les techniques utilisées dans cet axe.

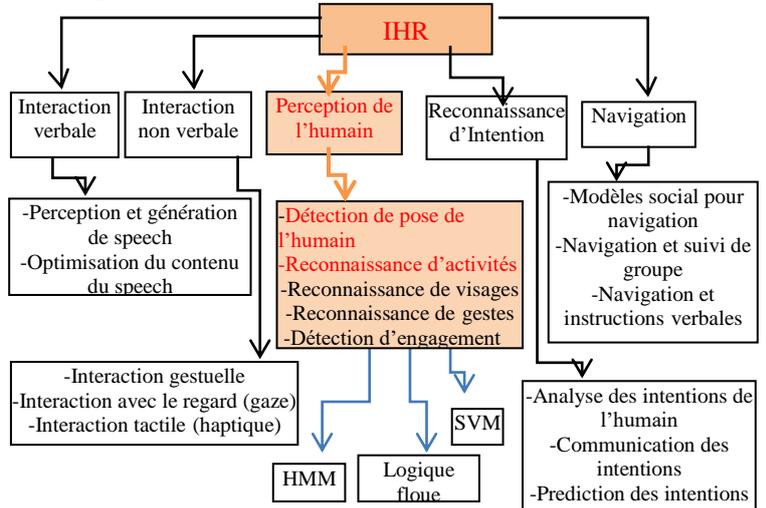


Figure 1 : Les principaux travaux de recherche dans l'IHR

II. PROBLÉMATIQUE

Depuis l'apparition de la camera RGB-D (KINECT) en 2010, équipée en plus, d'une camera RGB traditionnelle, d'une camera Infrarouge fournissant ainsi la profondeur de chaque pixel, plusieurs travaux ont été réalisés sur la reconnaissance de la pose de l'humain et de ses activités. Chungoo et al.[17] ont fourni une base de données de 18 activités contenant des gestes et des activités journalières de l'humain à partir d'une Kinect, leurs objectifs est de reconnaître les activités communicatives (interactives) et les activités non communicatives (non interactives).

Anjum et al. [18] ont utilisé un classifieur basé sur Support Vector Machine (SVM) pour classer les activités de l'humain à partir de la Kinect parmi un ensemble de 8 activités, pour chacune d'elles ils ont défini des articulations de référence à tracker. Leur méthode a donné un bon résultat de classification pour des activités clairement distinctes, mais moins précis pour des activités légèrement similaires.

C.Granata et al. [19] ont proposé une interface graphique offrant un certain nombre de services aux personnes âgées en les surveillant et analysant leurs activités pour cela ils ont implémenté un moteur de décision en utilisant la logique floue. Les entrées de ce moteur décisionnel sont des informations extraites de plusieurs capteurs et en sortie les valeurs d'intérêt associées à chaque action que le robot peut exécuter. Le robot peut alors décider dynamiquement quelle est l'action la plus adéquate à réaliser dans les différentes situations. L'engagement des personnes à rentrer dans une interaction avec le robot na pas été pris en compte dans ce travail, ce qui offrira plus de confort aux utilisateurs.

Plusieurs travaux sur la fusion multi-sensorielle basés sur les Modèles de Markov cachés HMM ont été proposés, dans le but de reconnaître les gestes et mouvements du corps humain

en temps réel [20] et [21]. Les inconvénients majeurs des approches basées HMM sont dû à : l'ajustement des probabilités qui est une tâche très difficile, leur efficacité dépend fortement de la quantité de détails fournis lors de l'apprentissage, temps de reconnaissance élevé, solution autour des maxima locaux [22].

III. TRAVAUX RÉALISÉS

Pour une interaction homme robot naturelle et intuitive, il est nécessaire pour le robot de détecter l'humain et sa pose. La détection de la pose permet au robot de savoir si l'humain se présente de face, de profil ou de dos (voir figure 2.a), cela lui permettra de s'assurer que l'humain s'apprête à engager une interaction avec lui. A cet effet, un travail a été réalisé dans l'objectif de calculer les angles d'orientation du corps humain à partir de la Kinect, et cela pour définir celles qui favorisent une interaction

Pour valider cette approche nous l'avons expérimenté sur le robot B21r du CDTA sous l'architecture robotique ROS dans deux applications ; la première pour le suivi d'un humain détecté et reconnu [23], et la deuxième dans un algorithme de navigation sociale [24] où le robot applique des règles sociales pour l'évitement des humains selon leurs poses (voir figure 2b).

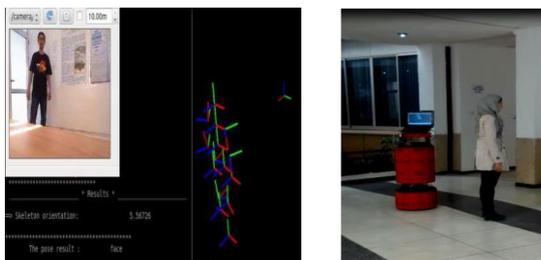


Figure 2 : (a) Détection de la pose de l'humain (b) Evitement par le Dos

IV. CONCLUSION ET PERSPECTIVES

Nous avons situé notre sujet dans l'état de l'art global du domaine d'IHR, et présenté brièvement quelques techniques utilisées pour la reconnaissance d'activités des humains. Nous avons aussi implémenté un algorithme de détection de pose de l'humain et calculé son orientation par rapport au robot à partir de la Kinect, ensuite nous avons validé l'algorithme dans un scénario de suivi et un autre de navigation sociale.

Les tâches suivantes restent à effectuer :

- Définir et créer une base de données personnalisée des activités adéquates et pertinentes d'un humain, permettant à un robot guide d'inférer ses besoins et d'interagir avec lui.
- Tester les algorithmes de reconnaissance d'activités (SVM et logique floue) sur notre base de données.
- Comparaison des résultats.

References

[1] C. Torrey, A. Powers, S. R. Fussell, and S. Kiesler. Exploring adaptive dialogue based on a robot's awareness of human gaze and task progress. In Proceedings of the ACM/IEEE international conference on Human-robot interaction, pages 247–254. ACM, 2007.

[2] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. A larger audience, please!: encouraging people to listen to a guide robot. In Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, pages 31–38. IEEE Press, 2010.

[3] M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski. Gracefully mitigating breakdowns in robotic services. In Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, pages 203–210. IEEE, 2010

[4] M. Salem, M. Ziadee, and M. Sakr. Effects of politeness and interaction context on perception and experience of hri. In Social Robotics, pages 531–541. Springer, 2013

[5] C. Perez Quintero, R. T. Fomena, A. Shademan, N. Wolleb, T. Dick, and M. Jagersand. Sepo: Selecting by pointing as an intuitive human-robot command interface. In Robotics and Automation (ICRA), 2013 IEEE International Conference on, pages 1166–1171. IEEE, 2013.

[6] J. Ham, R. H. Cuijpers, and J.-J. Cabibihan. Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. International Journal of Social Robotics, 7(4):479–487, 2015.

[7] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati. Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, pages 49–56. ACM, 2014.

[8] C. Bevan and D. Stanton Fraser. Shaking hands and cooperation in telepresent human-robot negotiation. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction pages 247–254. ACM, 2015.

[9] K. Talamadupula, G. Briggs, T. Chakraborti, M. Scheutz, and S. Kambhampati. Coordination in human-robot teams using mental modeling and plan recognition. In Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, pages 2957–2962.

[10] C.-M. Huang and A. L. Thomaz. Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In RO-MAN, 2011 IEEE, pages 65–71. IEEE, 2011.

[11] S. Satake, H. Iba, T. Kanda, M. Imai, and Y. M. Saiki. May i talk about other shops here?: Modeling territory and invasion in front of shops. In Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction, HRI '14, pages 487–494, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2658-2. URL

[12] R. Murakami, L. Y. Morales Saiki, S. Satake, T. Kanda, and H. Ishiguro. Destination unknown: Walking side-by-side without knowing the goal. In Proceedings of the 2014 ACM/IEEE International Conference on Humanrobot Interaction, HRI '14, pages 471–478, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2658-2. URL

[13] J. Brookshire. Person following using histograms of oriented gradients. International journal of social robotics, 2(2):137–146, 2010.

[14] H.-D. Yang, A.-Y. Park, S.-W. Lee, et al. Gesture spotting and recognition for human-robot interaction. Robotics, IEEE Transactions on, 23(2):256–270, 2007

[15] M. L. Anjum, O. Ahmad, S. Rosa, J. Yin, and B. Bona. Skeleton tracking based complex human activity recognition using kinect camera. In Social Robotics, pages 23–33. Springer, 2014.

[16] J. Lee, J. F. Kiser, A. F. Bobick, and A. L. Thomaz. Vision-based contingency detection. In Proceedings of the 6th international conference on Human robot interaction, pages 297–304. ACM, 2011.

[17] A. Chungoo, S. S. Manimaran, and B. Ravindran. Activity recognition for natural human robot interaction. In Social Robotics, pages 84–94. Springer, 2014.

[18] M. L. Anjum, O. Ahmad, S. Rosa, J. Yin, and B. Bona. Skeleton tracking based complex human activity recognition using kinect camera. In Social Robotics, pages 23–33. Springer, 2014.

[19] C. Granata, P. Bidand, A. Buendia, "Interactive person following for social robot: hybrid reasoning based on fuzzy and multiple-objectives decision making approaches", Int.Conf; on Climbing and Walking Robot CLAWAR 2011, France pp: 11-26.

[20] Sung, J., Ponce, C., Selman, B., Saxena, A. "Unstructured human activity detection from rgb-d images". In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 842–849 (2012)

[21] B. Burger, I. Ferrané, F. Lerasle, and G. Infantes. Two-handed gesture recognition and fusion with speech to command a robot. *Autonomous Robots*, 32(2):129–147, 2012

[22] Han Kuo, Chandimal Jayawardena "Application of Fuzzy techniques in Human-Robot Interaction- A review". In Social Robotics, pages 249–255. Springer, 2014.

[23] S.Kahlouche, N.Ouadah, M.Belhocine, M.Boukndoura, "Human pose recognition and tracking using RGB-D camera", 8th International Conference on Modelling, Identification and Control (ICMIC-2016), Algeria, pp 520-526. <https://doi.org/10.1109/ICMIC.2016.7804168>

[24] A.Bellarbi, S.Kahlouche, N.Achour, N.Ouadah, «A social planning and navigation for tour-guide robot in human environment», 8th International Conference on Modelling, Identification and Control (ICMIC-2016),Algeria, ,pp: 622 – 627, <https://doi.org/10.1109/ICMIC.2016.7804186>

Dr Mahmoud Belhocine :

Using background knowledge to enhance biomedical ontology matching

Amina ANNANE

Supervisors : Pr.Zohra Bellahsene, Dr.Faiçal Azouaou and Clement Jonquet

Résumé—Ontology alignment is the process of finding correspondences between ontologies. The use of external knowledge resources as background knowledge seems to be an effective solution to overcome the semantic heterogeneity. For that, we try to design a new approach that allows selecting (or constructing) dynamically from a set of knowledge resources a useful Background Knowledge for a given ontology matching task. Then, define how to use it effectively in the alignment process to obtain a high quality alignment.

I. INTRODUCTION

Ontology alignment is recognized by the scientific community as an important area of research because of its multiple applications in different domains [1]: ontology engineering, data integration, information sharing, etc. Especially in the biomedical domain that generates and manipulates a big volume of data. Ontology matching play a key role in the development of biomedical research by facilitating the development of data warehouses articulated around common ontologies. Many works have been made to extract mappings automatically, mainly using lexical and structural matchers [2], but these matchers often fail when the ontologies to align have different structures and do not use the same vocabulary (different terms to describe the same concepts)[3][4]. An alternative solution for automatic approaches is the use of external knowledge resources, commonly called background knowledge (BK), as a semantic mediator to discover mappings between ontologies (see Figure. 1). These background knowledge resources span from thesaurus, lexical resources, linked open data, one or several ontologies or a full repository of ontologies [5, 6, 7].

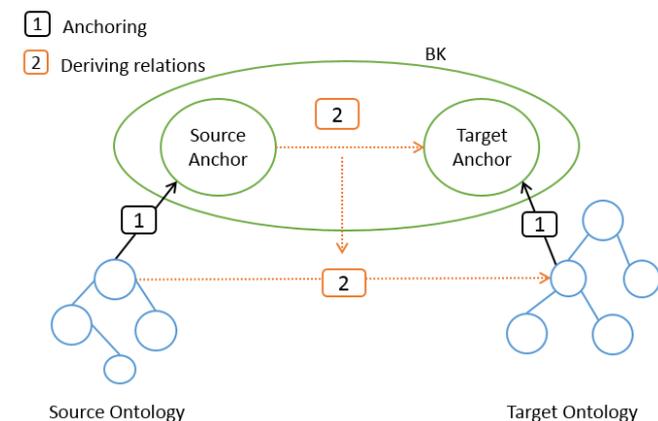


Fig 1. BK based matching

II. CHALLENGING ISSUES

The use of background knowledge in ontology alignment domain raises two main challenges: (1) its selection: How to select the most useful background for a given ontology matching task? (2) Its usage: How to use it, how to benefit from the knowledge of the selected BK to enhance alignment quality?

In our thesis, we deal with these two challenges.

III. RELEASED WORKS

A. Multilingual mapping reconciliation between English-French biomedical ontologies [8, 9]

The first work that we have done in the context of our thesis was the production of a resource that contains more than 228K mappings between French-English biomedical ontologies. Even if multilingual ontologies are now more common, for historical reasons, in the biomedical domain, many ontologies or terminologies have been translated from one natural language to another resulting in two potentially aligned ontologies but with their own specificity (e.g., format, developers, and versions). Most often, there is no formal representation of the translation links between translated ontologies and original ones and those mappings are not formally available as linked data. However, these mappings are very important for the interoperability and the integration of multilingual biomedical data. In this work, we proposed an approach to represent translation mappings between ontologies based on the NCBO BioPortal [15] format. We have reconciled more than 228K mappings between ten English ontologies hosted on NCBO BioPortal and their French translations. Then, we have stored both the translated ontologies and mappings on a French customized version of the platform, called the SIFR BioPortal [10], making the whole thing available in RDF. Reconciling the mappings turned more complex than expected because the translations are rarely exactly the same than the original ontologies. The multilingual mapping links produced in this study can have several applications including the integration of biomedical data of different languages, and multilingual semantic search and indexing.

B. Selection and Combination of Heterogeneous Mappings to Enhance Biomedical Ontology Matching [11]

In this work we proposed a novel BK based ontology matching approach that selects and combines dynamically existing mappings to build a background knowledge in graph form. The built BK is used to match ontologies. Current background knowledge approaches usually select either manually or automatically a limited number of different ontologies and use them as a whole for background

knowledge. Whereas in our approach, we propose to pick up only relevant concepts and relevant mappings linking these concepts from a preprocessed graph of existing mappings to build a specific background knowledge graph. Paths of this final graph will help to discover new mappings.

We have implemented and evaluated our approach using the NCBO BioPortal repository and the Anatomy benchmark from the Ontology Alignment Evaluation Initiative (OAEI). We have also used the mapping gain measure [14] to assess how much our final background knowledge graph improves results of state-of-the-art alignment systems. The evaluation shows that our approach produces a high quality alignment and discovers mappings that have not been found by state-of-the-art systems. The main contributions of this work are:

- A novel approach to align ontologies using a background knowledge graph automatically built from existing mappings.
- A novel measure to select the most accurate mappings from several candidate mappings derived from the previously built background knowledge graph.

C. Using Background Knowledge to Enhance Ontology Matching: a Survey

Several important questions related to the use of background knowledge resources arise (i) In which case is the use of BK resources justified and necessary? and (ii) What is the tradeoff between the complexity of the alignment methods and the use of BK resources in terms of the quality of matching and time execution? Another interesting issue is the selection of the useful BK resources for a given ontology matching task. In this work, we tried to answer these questions by reviewing the different approaches dealing with the two main steps of BK-based ontology matching that are BK resources selection and BK resources use. In addition, we provided a synthetic classification of BK resources use approaches. Finally, we present a comparative evaluation of BK-based ontology matching systems by analyzing their performance results obtained during Ontology Alignment Evaluation Initiative (OAEI) 2012-2016 campaigns. We thus evaluate the benefit of using BK resources and the improvement achieved by this approach with regard to the systems that do not use BK resources. This work will be submitted soon.

D. Combining Background Knowledge and machine learning to enhance ontology matching [13]

This work consists in refining and improving the idea originally presented in [11]. We proposed an approach to build a background knowledge containing only the right amount of concepts from a set of preselected ontologies, hence enhancing the efficiency of the matching process. The use of a background knowledge in ontology matching is a double-edged sword, it allows to increase recall (i.e., retrieve more correct mappings) but it may decrease precision (i.e., produce more incorrect mappings). To overcome this drawback, we propose a machine-learning based method to select the most relevant mappings from the candidates ones. Our experiments conducted on several Ontology Alignment Evaluation Initiative (OAEI) tasks show the effectiveness and efficiency of our approach. Our results are very competitive with the

state-of-the-art systems and overcome previously reported results in most tasks.

IV. PERSPECTIVES FOR THE NEXT YEAR

In addition to writing my thesis, we intend to investigate the use of BK in the multilingual context to match multilingual ontologies.

V. CONCLUSION

The use of BK in ontology matching is a promising solution. Our work confirms the benefit of using external resources, especially in the biomedical domain. However, not all domains have the richness of the biomedical domain in terms of the used BK resources (ontologies and terminologies). Consequently, more work should be done to investigate the use of generic BK resources such as DBpedia.

References

- [1] J. Euzenat and P. Shvaiko, *Ontology Matching*. Springer, 2013.
- [2] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, ‘Ontology matching: A literature review’, *Expert Syst. Appl.*, vol. 42, no. 2, pp. 949–971, 2015.
- [3] Z. Aleksovski, M. Klein, W. Ten Kate, and F. Van Harmelen, ‘Matching unstructured vocabularies using a background ontology’, in *International Conference on Knowledge Engineering and Knowledge Management*, 2006, vol. 4248, pp. 182–197.
- [4] P. Shvaiko and J. Euzenat, ‘Ontology matching: state of the art and future challenges’, *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 158–176, 2013.
- [5] C. Quix, P. Roy, and D. Kensché, ‘Automatic selection of background knowledge for ontology matching’, in *International Workshop on Semantic Web Information Management*, 2011, p. 5.
- [6] M. Sabou, M. d’Aquin, and E. Motta, ‘Using the semantic web as background knowledge for ontology mapping’, in *Proceedings of the 1st International Conference on Ontology Matching-Volume 225*, 2006, pp. 1–12.
- [7] J. Shamdasani, T. Hauer, P. Bloodsworth, A. Branson, M. Odeh, and R. McClatchey, ‘Semantic matching using the UMLS’, in *European Semantic Web Conference*, 2009, pp. 203–217.
- [8] Annane, A., Emonet, V., Azouaou, F., & Jonquet, C. (2016). Multilingual Mapping Reconciliation between English-French Biomedical Ontologies. 6th International Conference on Web Intelligence, Mining and Semantics, WIMS’16. pp:13:1-13:12.ACM.
- [9] Annane, A., Emonet, V., Azouaou, F., & Jonquet, C. Réconciliation d’alignements multilingues dans BioPortal. 27es Journées francophones d’Ingénierie des Connaissances, IC’16
- [10] Jonquet, C., Annane, A., Bouarech, K., Emonet, V., & Melzi, S. (2016). SIFR BioPortal: Un portail ouvert et générique d’ontologies et de terminologies biomédicales françaises au service de l’annotation sémantique. 16ème Journées Francophones d’Informatique Médicale, JFIM’16.
- [11] Annane, A., Bellahsene, Z., Azouaou, F., & Jonquet, C. Selection and Combination of Heterogeneous Mappings to Enhance Biomedical Ontology Matching. 20th International Conference on Knowledge Engineering and Knowledge Management, EKAW’16. pp:19-33.
- [12]
- [13] Annane, A., Bellahsene, Z., Azouaou, F., & Jonquet, C. Combining background knowledge and machine learning to enhance ontology matching. *Journal of Web Semantics* 2017 [Submitted].
- [14] D. Faria et al., ‘Automatic background knowledge selection for matching biomedical ontologies’, *PLoS One*, vol. 9, no. 11, p. e111226, 2014.
- [15] Noy Natalya F et al., BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37:170-173, 2009.

Un Modèle de Services Multifonctions pour les Systèmes Informatiques Ubiquitaires : Application au domaine Médical

Sihem BEY

Directeurs de thèse : Pr. Youssef AMGHAR, Dr. Fahima NADER

Résumé—Nous nous intéressons dans cette thèse à l'Internet des Objets (IoT), domaine relativement récent qui ouvre de nombreux champs pour la recherche. Il y avait 25 milliards d'objets connectés en 2015, il y en aura 50 milliards en 2020 selon une étude de CISCO. Les objets connectés sont tous les dispositifs que l'on peut connecter à Internet (smartphones, tablettes, capteurs, senseurs, ...). De nombreux dispositifs électroniques permettent ainsi de connecter le monde physique au monde réel. Dans ce travail de recherche, nous cherchons à modéliser la dimension ubiquitaire des systèmes d'information connectés à des dispositifs physiques. Au jour d'aujourd'hui nous avons, dans un premier temps, réalisé un état de l'art en mettant l'accent sur les solutions proposées pour l'architecture de l'IoT, et dans un deuxième temps identifié différents verrous et défis soulevés par ce domaine.

I. INTRODUCTION

Les objets de l'IoT (Internet of Things), sont équipés de capacités d'identification, de détection, de connexion et de traitement qui leurs permettent de communiquer entre eux, et avec les autres dispositifs physiques et les services sur internet en vue d'un même objectif [1]. L'IoT touche tous les domaines d'application : les transports et la santé sont probablement les domaines qui font le plus usage des technologies liées à l'IoT.

Dans cette thèse, nous cherchons, au sein d'une approche unifiée, à définir les principes fondamentaux pour la spécification des systèmes d'information connectés à des dispositifs physiques de façon ubiquitaire (anyone, anywhere et anytime). Nous proposons d'explorer de nouveaux modèles et outils qui permettent de prendre en compte l'interopérabilité au sein d'un système d'information connecté à des objets. Afin de donner un sens appliqué à notre recherche, nous nous intéresserons au domaine médical et en particulier au suivi des patients et des personnels hospitaliers ainsi qu'au domaine des urgences mobiles (ambulances équipées de dispositifs permettant de collecter des données du patient dès sa prise en charge).

II. MOTIVATION

Les objets intelligents participent d'un concept récent et promu à un bel avenir selon (Gartner et IDC [2, 3]). Il s'agit d'une nouvelle vision du futur de l'internet qui accueillera dans son réseau les objets que l'on trouve partout, dans les maisons, dans les villes (smart cities), les entreprises, les industries, les hôpitaux, etc.

L'IoT est présent de nos jours à travers son volet matériel avec des projets réellement aboutis et séduisants (Glycount : la

montre connectée pour surveiller le diabète [4], Tucky : le thermomètre pour surveiller la fièvre et la convulsion [5], etc.).

La propagation rapide de l'IoT et la quantité importante de données qu'il génère [6], suscitent l'intérêt de plusieurs chercheurs à travailler sur des modèles et des outils qui permettent la représentation, la collecte, l'interprétation, la transmission et l'analyse des données. De plus, l'IoT vise à rendre les objets 'intelligents' en les connectant non seulement au monde virtuel qui est l'internet et les applications, mais aussi entre eux. Rendre cette interconnexion tangible en dépit de l'hétérogénéité des objets, nécessite la mise en œuvre d'un modèle de communication qui permet une interopérabilité entre les objets.

III. ETAT DEL'ART

La revue de littérature [7, 8,9 ,10 ,11 ,12] que nous avons menée a montré qu'il existe plusieurs middlewares conçus pour assurer la communication entre les objets connectés et les applications. Un travail de classification et d'évaluation de ces middlewares a été effectué par des membres de l'école 'Computer Science and Statistics' [13] (IEEE 2016). Ils ont regroupé les middlewares selon leurs approches de conception en plusieurs groupes tels que : les middlewares à base d'événement, orienté service, à base de VM, à base d'agent, orienté base de données et orienté application.

La question de la communication entre les différents dispositifs de l'IoT et notamment la découverte automatique des ressources et des objets, a fait l'objet de nombreuses publications. De plus, plusieurs modèles de découverte orientés services, ont été proposés tels que ceux décrits dans [14, 15] ; toutefois en raison de la complexité du processus d'enregistrement des services dans les registres, et des ressources limitées des objets [16, 17, 18], les approches centralisées sont de plus en plus remises en cause. Il est clair que la recherche de solutions pour une meilleure gestion et traitement d'un volume très important de données, conduit les chercheurs à s'appuyer sur des infrastructures du Cloud Computing dans le monde de l'IoT [19, 20], compte tenu de ses ressources virtuelles illimitées en termes de stockage et de puissance de traitement [19].

Par ailleurs, des travaux de standardisation sont en cours de réalisation, tels que : IoT Forums du Worldwide qui travaillent pour développer un modèle commun qui assure une interopérabilité entre les objets intelligents, et Open Internet Consortium (OIC) qui se concentrent actuellement sur l'interopérabilité pour définir les spécifications et l'intégration

de milliards d'objets intelligents et les problèmes d'évolutivité [21].

Les technologies du web sémantique ont été utilisées dans plusieurs travaux [22-25], ces dernières années, pour répondre au problème de l'interopérabilité entre les dispositifs connectés hétérogènes. Ils proposent des modèles sémantiques pour la représentation et la manipulation des dispositifs de l'IoT. Par exemple, Ali Yachir et al [22] proposent un modèle sémantique pour la description des objets intelligents et la résolution des requêtes des utilisateurs à base d'ontologies. Ce modèle sémantique nécessite encore une évaluation empirique pour améliorer les classifications de services pour l'interopérabilité sémantique entre les objets intelligents. Sohail Jabbar et al [23] ont proposé le modèle IoT-SIM pour une interopérabilité sémantique entre les dispositifs de l'IoT. Ce modèle a été appliqué dans un domaine médical où les médecins peuvent surveiller leurs patients à distance.

IV. PROBLEMATIQUE

A travers cet état de l'art, nous constatons que le problème de l'interopérabilité nourrit la littérature. Il est considéré comme un vrai défi pour le développement à large échelle de l'IoT [23]. De nombreux travaux se sont intéressés à la communication entre les dispositifs physiques de l'IoT entre eux ainsi qu'avec des acteurs humains. Parmi ces travaux, on peut citer ceux qui proposent des middlewares comme Gateway entre le monde physique et le monde virtuel (internet) et ceux qui soutiennent l'idée de la standardisation en proposant la conception de modèles communs pour la représentation, la description et la découverte des dispositifs connectés.

Les middlewares proposés par la communauté répondent à de nombreuses exigences de l'IoT. Cependant, il reste encore des verrous à lever notamment pour répondre à des questions d'évolutivité, de fiabilité, de sensibilité au contexte, de sécurité et de vie privée [13].

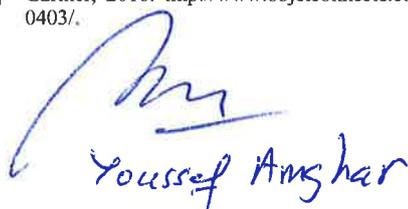
V. TRAVAIL EN COURS

Nous préparons actuellement un article de survey qui vise à résumer les travaux de recherche qui proposent des modèles de description sémantique pour la représentation et la gestion des dispositifs connectés dans l'IoT. Ce travail va nous permettre d'affiner notre problématique. Nous nous attacherons ensuite à développer, au sein d'une approche unifiée, l'ensemble des notions nécessaires à la définition des principaux mécanismes pour la conception de systèmes ubiquitaires dans le domaine de l'IoT.

Sur un plan pratique, nous envisageons d'évaluer notre approche dans le domaine médical. Nous chercherons dans la mesure du possible à nous approcher d'un centre hospitalier afin de rendre plus crédible notre domaine d'application.

Références

- [1] Bouguettaya, A., Singh, M., Huhns, M., Sheng, Q. Z., Dong, H., Yu, Q. & Ouzzani, M. (2017). A service computing manifesto: the next 10 years. *Communications of the ACM*, 60(4), 64-72.
- [2] Gartner, 2016. <http://www.objetconnecte.com/gartner-etude-societe-iot-0403/>.
- [3] IDC, 2016. http://lexpansion.lexpress.fr/high-tech/cinq-graphiques-pour-mesurer-le-boom-des-objets-connectes_1794945.html
- [4] Glycount. <http://webdesobjets.fr/objets-connectes/glycount-la-montre-connectee-pour-surveiller-son-diabete/>
- [5] Tucky. <https://www.aruco.com/2016/01/e-takescare-devoile-tucky/>
- [6] Pitney Bowes, 2016. <http://www.pitneybowes.com/fr/gestion-informations-client/case-studies/data-management-in-the-internet-of-things.html>
- [7] Pietzuch, P. R. (2004). *Hermes: A scalable event-based middleware* (Doctoral dissertation, University of Cambridge).
- [8] Eisenhauer, M., Rosengren, P., & Antolin, P. (2010). Hydra: A development platform for integrating wireless devices and sensors into ambient intelligence systems. In *The Internet of Things* (pp. 367-373). Springer New York.
- [9] Levis, P., & Culler, D. (2002). Maté: A tiny virtual machine for sensor networks. *ACM SIGPLAN Notices*, 37(10), 85-95.
- [10] Liu, T., & Martonosi, M. (2003, June). Impala: A middleware system for managing autonomic, parallel sensor systems. In *ACM SIGPLAN Notices* (Vol. 38, No. 10, pp. 107-118). ACM.
- [11] Murphy, A. L., Picco, G. P., & Roman, G. C. (2001, April). Lime: A middleware for physical and logical mobility. In *Distributed Computing Systems*, 2001. 21st International Conference on. (pp. 524-533). IEEE.
- [12] Han, Q., & Venkatasubramanian, N. (2001). Autosec: An integrated middleware framework for dynamic service brokering. *IEEE distributed systems online*, 2(7), 22-31.
- [13] Razzaque, M. A., Milojevic-Jevric, M., Palade, A., & Clarke, S. (2016). Middleware for internet of things: a survey. *IEEE Internet of Things Journal*, 3(1), 70-95.
- [14] Zambonelli F, Viroli M (2011) A survey on nature-inspired metaphors for pervasive service ecosystems. *Int J Pervasive Comput Commun* 7(3):186-204
- [15] Teixeira T, Hachem S, Issamy V, Georgantas N (2011) Service-oriented middleware for the internet of things: a perspective. In: Abramowicz W, Llorente I, SurrIDGE M, Zisman A, Vayssières J (eds) *Towards a service-based internet*. Springer, Berlin, Heidelberg, pp 220-229
- [16] Rapti, E., Karageorgos, A., Houstis, C., & Houstis, E. (2016). Decentralized service discovery and selection in Internet of Things applications based on artificial potential fields. *Service Oriented Computing and Applications*, 1-12.
- [17] He Q, Yan J, Yang Y, Kowalczyk R, Jin H (2013) A decentralized service discovery approach on peer-to-peer networks. *IEEE Trans Serv Comput* 6(1):64-75
- [18] He Q, Yan J, Yang Y, Kowalczyk R, Jin H (2008) Chord4s: a p2p-based decentralized service discovery approach. In: *IEEE international conference on services computing*, 2008. SCC'08, vol 1. IEEE, pp 221-228
- [19] Díaz, M., Martín, C., & Rubio, B. (2016). State-of-the-art, challenges, and open issues in the integration of internet of things and cloud computing. *Journal of Network and Computer Applications*, 67, 99-117.
- [20] Aazam, M., Khan, I., Alsaffar, A. A., & Huh, E. N. (2014, January). Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved. In *Applied Sciences and Technology (IBCAST)*, 2014 11th International Bhurban Conference on (pp. 414-419). IEEE.
- [21] <http://blogs.cisco.com/digital/iotmeets-standards-driving-interopability-and-adoption>.
- [22] Yachir, A., Djamaa, B., Mecheti, A., Amirat, Y., & Aissani, M. (2016). A comprehensive semantic model for smart object description and request resolution in the internet of things. *Procedia Computer Science*, 83, 147-154.
- [23] Ayari, N., Chibani, A., Amirat, Y., & Matson, E. T. (2015, September). A novel approach based on commonsense knowledge representation and reasoning in open world for intelligent ambient assisted living services. In *Intelligent Robots and Systems (IROS)*, 2015 IEEE/RSJ International Conference on (pp. 6007-6013). IEEE.
- [24] Albreshne, A., Lahcen, A. A., & Pasquier, J. (2014). Using a residential environment domain ontology for discovering and integrating smart objects in complex scenarios. *Procedia Computer Science*, 32, 997-1002.
- [25] Kovatsch, M., Hassan, Y. N., & Mayer, S. (2015, October). Practical semantics for the Internet of Things: Physical states, device mashups, and open questions. In *Internet of Things (IOT)*, 2015 5th International Conference on the (pp. 54-61). IEEE.



Youssef Amghar

Une approche d'aide à la constitution d'une source de données hétérogènes

Sabrina Abdellaoui

Directeurs de thèse : Fahima Nader

Équipe : MSI

Résumé—Les systèmes d'intégration des données (SID) permettent de combiner les données provenant de multiples sources et de fournir une représentation unifiée de ces données. L'intégration repose sur des processus d'Extraction-Transformation-Chargement (ETC), qui consistent à extraire, transformer, nettoyer des données issues de diverses sources pour les charger dans un entrepôt de données cible. La conception des processus ETC fait face à divers défis. Nous considérons dans cette thèse les problèmes majeurs liés à la qualité des données. Nous proposons une approche de conception ETC centrée sur la qualité de données.

I. INTRODUCTION

L'aide à la décision stratégique est confrontée à une situation inédite introduite par l'avènement de l'ère du big data. Les données sont générées, collectées et analysées à une échelle sans précédent afin de délivrer aux décideurs les informations pertinentes pour la prise de décision. Ces données sont issues de diverses sources, potentiellement externes à l'organisation. Les systèmes d'intégration des données (SID) permettent de combiner les données provenant de multiples sources et de fournir une représentation unifiée de ces données. On distingue deux types de SID selon la localisation des données: les SID virtuels et les SID matérialisés. Dans un contexte décisionnel, les SIDs matérialisés sont la solution la plus appropriée offrant divers outils d'analyse et de visualisation pour faciliter le processus de prise de décision. Un SID matérialisé repose sur des processus d'Extraction-Transformation-Chargement (ETC), qui consistent à extraire, transformer, nettoyer des données issues de diverses sources pour les charger dans un entrepôt de données (ED) cible. La conception des processus ETC est une tâche complexe qui fait face à divers défis:

1. La perpétuelle croissance du volume de données: avec l'avènement du big data nous assistons à la large prolifération de grand volume de données ainsi qu'à l'explosion du nombre de sources à prendre en considération par les processus ETC.

2. L'hétérogénéité de données: les types d'hétérogénéité qui peuvent affecter les sources de données peuvent être regroupés en deux grandes classes à savoir: (i) l'hétérogénéité structurelle qui est due à l'utilisation de différentes structures ou différents formats de stockage de données. (ii) l'hétérogénéité sémantique qui résulte des différences d'interprétation des objets du monde réel.

3. La qualité variable des données: les données considérées par les processus ETC sont de qualité très différentes. Des études récentes ont démontré que les problèmes liés à la qualité des données sont très répandus dans les grandes bases de données et sur le Web. Le recours à des données de qualité

médiocre ou incertaine menace la bonne prise de décision. Cette préoccupation augmente quand il s'agit de manipuler des données externes à l'organisation.

Nous considérons dans notre thèse les défis majeurs liés à la qualité des données lors de la conception des processus ETC. Les problèmes de qualité sont amplifiés par l'hétérogénéité et le volume des données à prendre en considération dans un contexte d'aide à la décision stratégique.

La qualité des données est un concept multidimensionnel qui implique plusieurs indicateurs tels que : la consistance, la précision, la fraîcheur, la complétude, etc. Elle peut être compromise par plusieurs facteurs, tels que les erreurs de saisie (par exemple, techer la place de teacher), présence de données incomplètes, diverses conventions et différents vocabulaires pour représenter et stocker les objets du monde réel (par exemple, 13 Fifth Street Ouest contre 13 W. 5th St.), absence ou manque de contraintes d'intégrité.

II. PROBLÉMATIQUE

Pour assurer la qualité des données, une variété de règles de qualité de données traditionnelles ont été utilisées pour capturer les incohérences de données et les réparer telles que: Dépendances fonctionnelles (DF), dépendances fonctionnelles conditionnelles (DFC), Denial Constraints (DCs) et les dépendances de correspondances (DM). Cependant, peu d'approches ont porté sur le nettoyage des données dans le contexte de la conception de la phase ETC. Les approches traditionnelles d'ETC se sont concentrées sur la génération des correspondances entre le schéma des sources de données et l'ED cible, et la réduction du temps d'exécution des processus ETC. Malgré le fait que ces processus sont censés garantir un nettoyage approprié des données, la détection et la réparation des incohérences restent principalement basées sur des règles et des programmes de bas niveau, ce qui nécessite une implication importante des utilisateurs. Par conséquent, l'intégration du nettoyage des données dans la conception de la phase ETC est une question cruciale qui doit permettre:

- A) La détermination des règles de qualité de données et des besoins en qualité liés à un contexte déterminé.
- B) La représentation de plusieurs types de règles de qualité de données d'une manière uniforme et flexible au sein de l'ED.
- C) L'identification des problèmes de qualité de données comme étant des violations des règles spécifiées.
- D) L'évaluation de la qualité des données.
- E) La correction des problèmes de qualité identifiés afin de peupler l'ED avec des données de bonne qualité.

III. TRAVAUX RÉALISÉS

- A) La proposition d'une approche ontologique de conception d'un ETC dans un contexte d'aide à la décision stratégique, cas d'application: Intelligence économique [1], [2].
- B) La proposition d'un métamodèle dédié à la qualité de données [4].
- C) La proposition d'une approche pour la conception d'un ETC centrée sur la gestion de la qualité de données ontologiques [3].
- D) La prise en compte des besoins en qualité durant le processus de conception de la phase ETC par: (1) La sélection des données critiques concernées par le processus d'évaluation de la qualité, (2) La détermination des critères de qualité pertinents selon les experts de domaine et les utilisateurs, (3) La définition des règles de qualité, (4) La liaison entre chaque règle et son critère de qualité correspondant, (5) L'affectation d'un seuil de qualité attendu par les experts de domaine et les utilisateurs, relatif à chaque critère de qualité, (6) La détermination des métriques pour mesurer les critères de qualité de données. Ces métriques sont élaborées pour un contexte d'évaluation spécifique [4].
- E) L'extension du modèle de l'ontologie de l'ED cible pour permettre la représentation des règles de qualité de données. Les inconsistances sont ainsi identifiées comme étant des violations des règles spécifiées [5].
- F) L'évaluation de la qualité des données :Cela implique le calcul des scores des métriques associées aux critères de qualité et de les comparer aux seuils attendus déterminés par les experts de domaine et les utilisateurs et déterminer de ce fait si ces données répondent aux besoins en qualité définis [4].
- G) La proposition d'une nouvelle méthode de réparation des données qui corrige automatiquement les données erronées lorsque cela est possible ou génère des réparations possibles. Les utilisateurs sont impliqués dans le processus de réparation pour valider les réparations possibles retournées ou suggérer de nouvelles réparations [5].
- H) La réalisation d'un outil pour concrétiser l'approche proposée ainsi que des expérimentations approfondies pour démontrer la généralité, l'efficacité et l'efficacité de notre approche en utilisant des ensembles de données synthétiques et réelles [5].

IV. PERSPECTIVES

Nous travaillons actuellement sur la rédaction de la thèse.

V. CONCLUSION

Nous avons présenté dans ce document, un résumé de l'état d'avancement de notre thèse de doctorat. Les principaux travaux réalisés ainsi que les travaux futurs.

Nous proposons une approche de conception d'un ETC centrée sur la qualité des données dans un contexte d'aide à la décision stratégique qui lève les verrous suivants :

Comment identifier et résoudre l'hétérogénéité des données présentes dans les différentes sources ?

Quelle est la structure générique à adopter pour l'intégration des données ?

Comment identifier les données inconsistantes et les erreurs ?

Comment assurer que les données à intégrer dans l'entrepôt de données répondent aux besoins en qualité ?

Comment réparer les inconsistances identifiées ?

References

- [1] S. ABDELLAOUI, F. NADER, « Une approche d'intégration pour la conception d'une source de connaissance dédiée à l'intelligence économique, » Séminaire International sur l'Intelligence Economique, Alger, Mai 2014.
- [2] S. ABDELLAOUI, F. NADER, « Semantic Data Warehouse at the heart of Competitive Intelligence Systems: design approach, » 6th. edition of the International Conference in Information Systems and Economic Intelligence(SIIE 2015), Hammamet, Tunisia, 2015.
- [3] S.KHOURI, S. ABDELLAOUI, F. NADER, «Avoiding Ontology Confusion in ETL Processes, » 19th East-European Conference on Advances in Databases and Information Systems(ADBIS 2015), Futuroscope, Poitiers-France, September 2015.
- [4] S. ABDELLAOUI, L.BELLATRECHE, F. NADER, «A Quality-driven Approach for Building Heterogeneous Distributed Databases: The Case of Data Warehouses, » Cluster, Cloud and Grid Computing (CCGrid), 2016 16th IEEE/ACM International Symposium. IEEE, 2016.
- [5] S. ABDELLAOUI, F. NADER, R. CHALAL, « QDflows: A System Driven by Knowledge Bases for Designing Quality-Aware Data flows., » Journal of Data and Information Quality (JDIQ), 8(3-4), 14, 2017.