



Université Hassiba BEN BOUALI de Chlef
Faculté des sciences
Département d'informatique

Proceedings ICNTC'2015



2^{ème} Conférence Internationale sur les Nouvelles Technologies et la Communication

Chlef, Algérie

3-4 mars 2015

<http://www.univ-chlef.dz/ICNTC/ICNTC2014>



***Second International Conference on New Technologies
and Communication –ICNTC’15-
03rd/04th March 2015
Hassiba Benbouali University of Chlef, Algeria***

OVERVIEW

ICNTC’15 Is the second International Conference on New Technologies and Communication that will be held from 3 to 4 march 2015 at Hassiba Ben Bouali university of Chlef, Algeria.

ICNTC’15 aims at fostering advances in the areas of New Technologies and Communication, protocols, Wireless Networks, Mobile Computing, Ad hoc and Ambient Networks, QoS, Web services, Web mining, distributed systems, Internet of Things, wireless networking for moving objects etc.

ICNTC’15 will address all the challenges of building the Future of communication and internet, which will be based on mobile, wireless and fixed broadband communications infrastructures. ICNTC’15 will become a stage for researchers, officials, decision makers, practitioners from the public sector and industry, people involved in other related initiatives and many others. The Conference will facilitate knowledge transfer to the community and practitioners by promoting results dissemination.

ICNTC’15 provides a dynamic forum for researchers, students and professionals to present their research works and development in these interesting areas.

Themes

Topics of the conference include, but not limited to, the following research areas:

- Wireless sensor networks, Ad hoc Networks, Mobility and security.
- Routing, QoS and scheduling protocols.
- Communication protocols.
- Computer Security.
- Cloud computing.
- Distributed systems.
- Grid Computing.
- Web Commerce & Services, Data models, Web searching & querying.
- Web Mining, Web services & Web Semantics.
- Web service based Grid computing and P2P computing.
- Mobile/wireless content distribution.
- Signal Processing.
- Natural language processing.
- E-learning and m-learning.
- Mobile Social Networks
- Peer-to-Peer Social Networks
- Ontologies
- Image processing
- Pattern recognition
- Data mining
- Radio Frequency and Antenna design
- Optical communications and devices
- Signal processing techniques for Communications

Conference Honor chairman

Professor: **Berrabah BENDOUKHA** Chlef University Rector

Conference General Chairman

Professor: **Abdelkader ALI BENAMARA** Science Faculty Dean

TECHNICAL PROGRAM COMMITTEE

Abassia DEBA	Oran University (Algeria)
Abdelkader BENYETTOU	USTO University (Algeria)
Ahmed LEHIRECHE	SBA University (Algeria)
Ahmed GUESSOUM	USTHB (Algeria)
Benamer KADRI	Tlemcen University (Algeria)
Bouabdellah KECHAR	Oran University (Algeria)
Belabbas YAGOUBI	Oran University (Algeria)
Bernard COUSIN	IRISA of Rennes (France)
Bouziane BELDJILLALI	Oran University (Algeria)
Congduc PHAM	Pau University (France)
Danda B. RAWAT	Georgia Southern, (USA)
Djamel-Eddine ZEGOUR	ESI (Algeria)
Fatima BENDELLA	USTO (Algeria)
Fatima DEBBAT	Mascara University (Algeria)
Fatima Zohra BELKREDIM	Chlef University de (Algeria)
Ghita KOUADRI MOSTEFAOUI	Diamond Light Source (Oxford,UK)
Hafidh HAFAF	Oran University (Algeria)
Khadija YAHYAOUI	Mascara University (Algeria)
Ladjel BELLATRECHE	(France)
Larbi SEKHRI	Oran University (Algeria)
Madaiah KRISHNAMURTHY	Bangalore, (India)
Mohamed BENMOHAMMED	Constantine University (Algeria)
Mounira OUARZEDDINE	USTHB, (Algeria)
Mohamed BENYETTOU	USTO University (Algeria)
Mohamed ELHASSOUNI	Rabat (Morocco)
Mohammed FEHAM	Tlemcen University (Algeria)
Mohamed SENOUCI	Oran University (Algeria)
Mahfoud BENCHAIBA	USTHB University (Algeria)
Mourad GHASSAN	Beyrouth (Liban)
Nabiha AZIZI	Annaba (Algeria)
Narimene BOUSTIA	Blida (Algeria)
Riad MOKADEM	Toulouse (France)
Safia Nait BAHLOUL	Oran University (Algeria)
Sidi Mohamed SENOUCI	Bourgogne University (France)
Tewfiq EL MALIKI	Genève(Switzerland)
Walid Khaled HIDOUCI	ESI (Algeria)
Yamine AIT AMEUR	ENSMA (France)
Youssef FAKHRI	Kenitra (Morocco)
Zakaria MAAMAR	Dubai (AEU)
Zohra SLAMA	Sidi Belabess University (Algeria)
Zoulikha MEKKAKIA MAAZA	USTO university (Algeria)

ORGANIZING COMMITTEE

Organization committee honor president

Professor: BELMADANI Bachir .(Vice-rector of scientific researches)

President

Ahmed LOUAZANI

Members

Ahmed HARBOUCHE .1	9. Sarah IBRI
Abdellah KOUIDER ELOUHED .2	10. Nassim DENNOUNI
Lahcen GUEMIDI .3	11. Mounir TAHAR ABBES
Mohamed SLIMANE .4	12. Rachid BECHAR
Mohamed ARIDJ .5	13. Mourad LOUKAM
Abdelmadjid ALLALI .6	14. Djelloul BETTACHE
M'Hamed HADJ HENNI .7	15. M'Hamed ZOUATNIA
Djamila HAMOUCHE .8	16. Ahmed SLIMANI

Contacts

Phones : +213 27 72 10 67/ 68 -- +213 27 72 33 83

Fax: +213 27 72 17 88

Email : icntc2014@gmail.com

Adresse : ICNTC'15 Conference
Département d'informatique, BP 151
Université Hassiba Benbouali de Chlef
Hay Salem, route nationale N° 19
02000 Chlef, Algérie

Web site : <http://www.univ-chlef.dz/ICNTC/ICNTC2014>

Front Page design: Nassim DENNOUNI & Rachid BECHAR

Edited by : Ahmed LOUAZANI & Nassim DENNOUNI

C o n t e n t s**P l e n a r y**

Ingenierie De La Connaissance -----	5
Polarimetric, multifrequency sar radars: a new era for remote sensing -----	5
Quantum algorithms and quantum computers a new idea to strength computational security. -----	6

P r e s e n t a t i o n

The evolution of collaborative tagging systems and the relation with the emergence of language: Agent based modeling -----	7
Deployment approaches of wireless sensor networks considering lifetime extension -----	12
Proposition D'un Protocole De Communication Dans Un Réseau Hybride Pour Améliorer La Sécurité Routière -----	17
Un protocole de négociation pour l'ordonnancement distribué Multi-Agents d'un Atelier Job-Shop -----	25
Coordination based agents for path finding -----	32
Replication Strategies in a Mobile Environment -----	38
Comparaisons des approches d'interopérabilité des réseaux P2P -----	45
Intégration d'un Module Morpho-lexical à partir du système Nooj dans une Plateforme d'Analyse de Textes en Arabe standard -----	53
Mobility Models for VANET simulation -----	62
New Half-mode Waveguide Designs based on SIW Technology for X band Applications -----	66
Stockage distribué de données dans les réseaux de capteurs sans fil -----	71
Bi-Clustering Algorithm Using Formal Concept Analysis -----	78
Simuler pour Evaluer les BMOs: Quelle Méthode Choisir ? -----	84
Exploration diversifiée par entités nommées -----	91
Une approche sémantique pour la recommandation de services dans le cadre d'une formation en ligne --	97
A Topology Management scheme with scalability and QoS guarantee for Large Scale Mobile Ad Hoc Networks in Urban Environment -----	104
Overview of Incentive Mechanisms in Peer-to-Peer Systems -----	110
Optimisation du routage des navires par une approche hybride basée sur l'ACF -----	118
TRACIS: a system for interpreting learners' traces -----	124
Towards Building Web Service Ontology: A WSDL Reverse Engineering Approach -----	130

P o s t e r s

CS-DSM : Cloud Service Description and Selection Model -----	137
ACOHM : Bio-inspired routing protocol for High Mobile Ad Hoc Networks -----	137
Routage sémantique d'information dans un réseau p2p de partage de fichier -----	137
The Fusion of SVD and NLS inversions for the Retrieval of a Simulated 3D Urban Structures in a Tomography SAR Application -----	138
Les jointures des flux de données dans les réseaux de capteurs sans fil -----	138
Parallel Pipelined Implementation of DES Cryptographic Algorithm on Multicore Machines -----	139
Election de Leader dans les Réseaux Mobiles Ad Hoc Basée sur le Protocole de Routage TORA -----	139
Extraction des relations de causalité dans le domaine médical -----	140
Vers une simulation du système d'orchestration des activités mobiles d'apprentissage à l'aide des SMAs	140

Ingenierie de la Connaissance

LEHIRECHE Ahmed

Laboratoire EEDIS, Université DJILLALI LIABES De Sidi Bel Abbès.

elhir@univ-sba.dz

Abstract- L'ingénierie des connaissances est une spécialité qui s'occupe de ce qui est extraction, formalisation et maintenance de la connaissance. Dans le domaine de l'informatique elle est utilisée pour ajouter la dimension sémantique aux technologies de l'information via les Systèmes à base de connaissances ou toute autres outils de manipulation de la connaissance. L'extraction consiste à utiliser des corpus (ex. Linguistique) pour déterminer le sens d'une unité lexicale (ex. model espace de mots). Les outils formels pour la spécification, la description, et l'inférence cognitive sont au centre du domaine IC ; tel les logiques de descriptions (DL), les Graphes conceptuels(CG) et les frames.

Sur un exemple réel ; Dans cette exposé nous présenterons comment utiliser les DL pour spécifier la connaissance, Comment penser en DL et Enfin quelques thèmes de recherches.

Mots Clés : Connaissance, Formalisme, DL, CG, Frame, Linguistique computationnelle, les Ressources, Corpus, Ontologie, TALN, Web sémantique, Métriques.



Biographie :

Ahmed LEHIRECHE has completed respectively ING Diploma from ESI of Algiers (1981) with the final curriculumproject at the IMAG(France), "MAGISTER" Diploma from USTOran (1993) and "DOCTORAT D'ETAT" Diploma from UDL Sidi bel Abbès (2005).

He is working as a Director of research, head of the Knowledge Engineering Teamat the EEDIS laboratory and full Professor at the computer science departmentof UDL Sidi bel Abbès. He is mainly concerned with AI, Computer Science Theory and Semantics in IT.

Polarimetric, multifrequency SAR radars: A new era for remote sensing

Abstract- Remote Sensing is the art of acquiring information about the Earth's surface without actually being in contact with it. This is done by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information. Today, the use of passive as well as active sensors is becoming more and more necessary for environment monitoring. The ability of active sensors such as SAR radar to acquire images day and night and in all weather conditions gives this type of sensor the advantage to be the unique solution in case passive sensors fail to observe the earth and with less cost. The advent of multipolarization and multifrequency radars enters the radar SAR to a new era, by allowing a continuous surveillance of all types of land use/ land cover areas. Applications of SAR radar multipolarized/multifrequency are among others: agriculture using multitemporal information, Land cover classification, Vegetation parameter retrieval, Soil moisture, snow depth, DEM generation and Biomass and height estimation.

Key words; Remote sensing, Radar, polarimetry, multifrequency, SAR

Biographie: Mounira Ouarzeddine is a member of the LTIR laboratory (Laboratoire de Traitement d'Images etRayonnement). She graduated in electronic engineering in 1993 at the university of Science and technology Houari Boumedienne (USTHB), in 1998 she finalized her magistère in electronics systems at the same university. In 2002 she got her Master of Science in Geoinformatics from ITC, Twente university, the Netherlands, Then back to USTHB, Algiers, she achieved her doctorat d'Etat in Image processing and remote sensing in 2008. Her main interests are signal processing applied in SAR radar, Interferometry, tomography and polarimetry.

Quantum Algorithms and quantum computers a new idea to strength computational security.

Abdellah Oumsalem

The Laboratory for Theoretical Physics and Material Physics -Chlef University- Chlef
 Preparatory school of sciences and techniques
 El Emir Abdelkader Bab El Oued, Algeria
abdnab2003@gmail.com

Abstract -During the past forty years astounding advances have been made in the field of quantum information theory. Many reasons lead to this development. First, Components shrink to where their behavior will soon be dominated more by quantum physics. Second, the physical limitations of the classical computer. Third, the characteristics offer by the exploitation of the quantum information theory. Velocity is one of these characteristics. As an example, a quantum algorithm can find such an item in a time proportional to the square root of the size of the set, which is considerably faster than classical methods that take the same time as the size of the set. In this work we have to show how a quantum algorithm is faster than a classical algorithm. As a case of study we compare between the Classical Search Algorithm (CSA) and the Grover's Algorithm (G'sA).

In computer science, CSA is a search algorithm which is suitable for searching a set of data for a particular value. It operates by checking every element of a list one at time in sequence until a match is found. The pseudo code describes the CSA is: «For each item in the list check to see if the item you're looking for matches the item in the list. If it matches return the location where you found it. If it does not match continue searching until you reach the end of the list. If we get to the last element, we know that the item does not exist in the list». As a result, the algorithm operates by checking every element of a list one at time in sequence until a match is found. CSA runs in O(N). If the data are distributed randomly, on average (N+1)/2 comparisons will needed. The best case is that the value is equal to the first element tested, in which case only one comparison is needed. The worst case is that the value is not in the list or in the last item in the list, in which case N comparisons are needed.

Otherwise, Grover's Algorithm runs $O(\sqrt{N})$. G'sA perform the search quadratically faster than can be done classically. This due to the use of the quantum superpositions and the state amplitude. We applicate G'sA as follow: we begin with lunch the n-qubit state $|00000 \dots 0\rangle$. After, we prepare the state

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle$$

by the application of the n-qubit Hadamard gate (H). As a final step, we apply the Grover

iterate (G) as $\left[\frac{\pi}{4} \frac{1}{\sqrt{N}} \right]$ times and we measure the resulting state. In these procedures Hadamard gate and Grover iterate work to increase the probability amplitude of the searching state $|I \varphi\rangle$. As a result we compare the amplitude of all states with the $|I \varphi\rangle$ amplitude. This gives us zero probability with all cases. However, it gives us one with just one element which is the searching element. This happens in \sqrt{N} time.

Finally, comparing to CSA which takes N time to search an appropriate element in a list of data, G'sA does a generic search using amplitude comparison techniques to find the same element in \sqrt{N} time. Also, G'sA is a good example to show how a quantum algorithm is powerful.

REFERENCES

- [1] M. A. Nielsen. Quantum Information Theory. Ph.D thesis, University of New Mexico, 1998.
- [2] M. Mosca and A. Ekert. The hidden subgroup problem and eigenvalue estimation on a quantum computer. arXiv e-print quant-ph/9903071, 1999.
- [3] Y. I. Manin. Classical computing, quantum computing, and Shor's factoring algorithm. arXiv e-print quantph/ 9903008, 1999.
- [4] X. Zhou, D.W. Leung, and I. L. Chuang. Quantum logic gate constructions with one-bit "teleportation". arXiv e-print quant-ph/0002039, 2000.
- [5] P.W. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. arXiv e-print quantph/0003004, 2000.



Biography:

Abdellah Oumsalem is an Assistant Professor of Electronics and scientific English at the Preparatory school of Sciences and Techniques, Algiers- Algeria. His main interests are "Nanotechnology and quantum algorithms".

The evolution of collaborative tagging systems and the relation with the emergence of language: Agent based modeling

Ahmed Mokhtari

École nationale Supérieure d'Informatique (ESI ex. INI)
Algiers, Algeria
ad_mokhtari@esi.dz

Salima Hassas

University Claude Bernard - Lyon 1
Lyon, France
hassas@univ-lyon1.fr

Abstract—In social web and in the collaborative tagging system in particular, users annotate resources using key words (tags) in order to categorize and organize their personal resources. In this paper, we study the emergence of tagging language formed by (tag/resource) couples shared by the users of the system. Using agent-based modeling, following the perspective of complex system, we show that a shared language can emerge in a collaborative tagging system.

Keywords—social web, collaborative tagging systems, language emergence, collaborative tagging, tag, Multi agents systems.

I. INTRODUCTION

In social web and in the collaborative tagging system in particular, users annotate resources by tags to create a categorization system. To what extent, the proposed tags help the emergence of a tagging language shared by the users of the collaborative tagging system?

Our motivation to this work is to improve the existing tagging systems by an emergent language that is established and will be used by the users in describing resources on the web and also will be used in their search in the system.

The works done in the modeling of the language emergence and language evolution have shown that group of agents can establish a common language, by applying certain principals like self-organization, imitation and reinforcement learning[1]. These principals help in the emergence and the evolution of a shared language.

In this paper, our contribution is the study of cognitive (linguistic) aspect of the tagging systems by applying emergence of language principals on collaborative tagging system where the users are modeled and simulated by multi agents system and showing that the users of such system develop a shared language that can be used on the search of the resources.

Related work

Among the works that are close to our work which is the evolution of collaborative tagging systems and the emergence of language, we cite as examples:

The work presented in [2] is concerned with the structure of collaborative tagging systems and the dynamics observed in these systems and it shows that the growth of the number

of the tags is described by a power law. In [3], there is a study of the evolution of tagging systems within a semiotic perspective; in [4] an epistemic model for collaborative tagging system is presented.

A study of the complex dynamics of tagging systems is presented in [5] where tripartite model is introduced for modeling the structure of the tagging system. In the work of Santos-Neto and Condon [6] individual and social behaviors are analyzed in tagging systems and probabilistic model is presented.

In the works of Steels and Kaplan [1] [7] [8] [9] [10] emergence language model is developed in which group of agents establish a common language and it can be considered as good framework for studying the emergence and evolution of language.

A. Problematic

This article aims to study the evolution of collaborative tagging systems and the conditions for the emergence of folksonomies, and also we present a study of these systems in a linguistic perspective considering folksonomies as language to answer the following question: Is there a relationship between the emergence of language and the evolution of tagging systems? What are the linguistic foundations that can explain the evolution of these systems to the emergence of folksonomies?

This problematic is addressed by using multi-agents based modeling.

B. The tripartite structure of tagging systems

The tripartite model has been theorized in [5]. There is three main entities that form any collaborative tagging system:

- The system's users.
- The tags themselves.
- The resources being labeled.

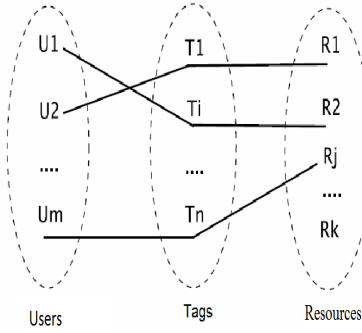


Fig. 1 - The tripartite structure

Each of the three entities may be seen as separate spaces which form sets of nodes connected by edges.

C. Definition of folksonomy

A folksonomy is a tuple $F = (U, T, R, Y, pt)$ [4] where

- $U, T,$ and R are finite sets, whose elements are called users, tags and resources respectively.
- Y is a ternary relation between them, $Y \subseteq U \times T \times R$, which represent the action of assigning tag.
- Pt is a function: $Y \rightarrow n$ which assigns to each element of Y timestamp n . It corresponds to the time when a user has assigned a tag to the resource.

D. The evolution and dynamics of a tagging system

The study of collaborative tagging systems shows that they evolve over time by complex dynamics [2]. This dynamics is held by the user-system interactions through resources annotation. The observation of the use and the reuse of tags shows that the growth of the number of the tags is described by a power law with an exponent smaller than one and this distribution gives rise to emergence of folksonomies (Fig.2).

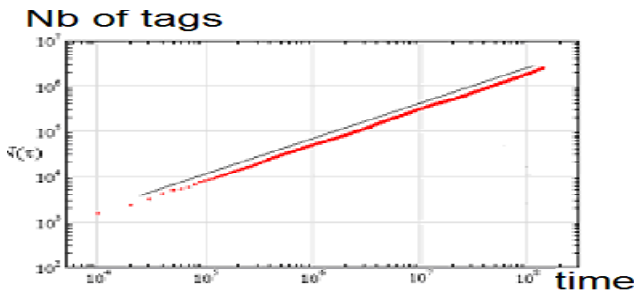


Fig. 2 Growth of the number of tags in a real tagging system

II. MODELING DYNAMICS OF TAGGING SYSTEM

For this model, our aim is to generate the dynamics observed in the collaborative tagging system and in particular the evolution of the number of tags that follow a "power law" distribution (fig.2).

The values of the parameters of the model are taking from the work of Santos-Neto and Condon [6] where the individual and social behaviors in tagging systems are analyzed and modeled analytically but we use agent based modeling rather than analytical model.

A. Model components

A collaborative tagging system consists of three parts, a set of users, a set of tags and a set of resources. These three sets are connected by links made by operation of collaborative tagging. In our model:

- System users are simulated by software agents,
- Tags are randomly generated words.
- Resources are represented by keywords also randomly generated.

B. Model parameters

The most important parameters of our model are:

- The number of the agents that access the system and participate in the tagging operation.
- The number of the most used tags presented to an agent to allow a social imitation.
- The threshold for measuring the similarity between the resources, it used to assign resources that are close to the same tag, for the similarity measures between two resources r_i, r_j represented by the two keywords key_word_i and key_word_j respectively, we used a lexical function defined as

$$Sim(r_i, r_j) = \frac{2 * Communs_prefix(key_word_i, key_word_j)}{Length(key_word_i) + Length(key_word_j)}$$

C. Interaction

The interaction of the agents with the system is as follows:

1. An agent accesses the system.
2. The system provides access for that user to all tags used in the previous sessions; the system also provides access to a set of most used tags by other users.
3. The agent chooses a new resource or an existing resource in the system.
4. The agent has three options to choose a tag to be assigned to the selected resource:
 - He chooses a tag from his personal tags; in this case we have a self imitation.
 - He chooses a tag amongst most used ones by the community of the tagging system users; in this case we have a social imitation.
 - He creates a new tag if he determines that the resource belongs to a new category that does not exist in its personomy (personal tags) nor in the folksonomy.
5. He sends the identifier of the resource and the tag he chose to the system as an input " U_i, T_i, R_i, t_i " Where U_i, T_i, R_i, t_i represent the user, the tag the resource, a time stamp.

D. Results and Discussion

We used Jade framework and java programming language in the development of our multi agents system.

We will present some results of a simulator designed based on the model described above. Our goal in this phase is to regenerate the complex dynamics of tagging systems. By modifying the parameters of our model, we obtain the same dynamics observed in the real collaborative tagging systems.

Example 1

In the first simulations, we use a very limited number of agents as for example $N = 10$ and a threshold agents similarity measure $S = 0.4$. The figure below shows the evolution of the number of tags over time.

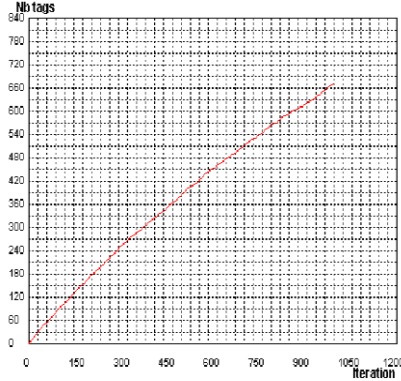


fig. 3 -Evolution of the number of tags

In the figure 3, the convergence of the number of tags to 671 tags can be observed after 1019 iterations and it's the same dynamics observed in a real collaborative tagging system (a power law).

Example 2

In this example, we use a number of agents $N = 500$ agents and a threshold of similarity measure $S = 0.4$.

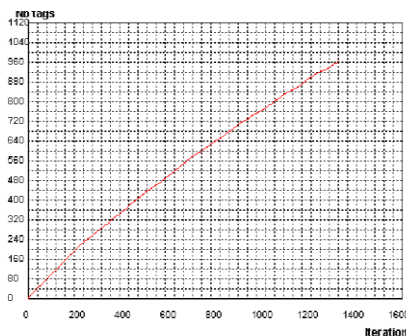


Fig. 4 -Evolution of the number of tags

In the figure 4, The convergence of the number of tags to 962 tags can be observed after 1363 iterations and it's the same dynamics observed in previous example and in the tagging system.

This is due to the principle of individual and social imitation used in our model which is based on the reuse of system's tags that allows the convergence and the emergence of folksonomy.

E. Summary and Discussion

From the examples above and from other examples, we see the emergence of folksonomies in the simulated tagging

system and the generation of the dynamics of such systems. The quality of the folksonomy in terms of number of tags depends on the threshold chosen for defining inter-resource similarity which defines somehow the degree of freedom for the agent to consider two resources as having the same content so its must have the same tag.

III. LINGUISTIC MODELING OF TAGGING SYSTEM

We are interested in this model to the cognitive and especially to the linguistic aspect of the process of collaborative tagging.

To study the linguistic aspect of the tagging process, we enrich the previous model as follows:

- We endow the agents with cognitive structures (associative memories); these memories aim to store linguistic knowledge about the assignments of tags to resources (couples tag / resource).
- A reinforcement learning mechanism so that agents can learn and adapt their knowledge by updating their associative memories.
- The agents interact with the system by following scenarios of language games [1].

This model is based in its linguistic side on the models proposed in the work of Kaplan [1].

A. Interaction

The novelty in the interactions compared to the previous model is that they follow schemes inspired of language games adapted for collaborative tagging systems because the interactions occur between an agent and the system (not between two agents like in standard language games) where the agent (resp. system) plays the role of speaker or interlocutor to negotiate the assignment of a word (a tag) to a meaning (a resource). The interaction of the agents with the system is as follows:

1. An agent accesses the system.
2. The system provides access for that user to all tags that used in the previous sessions; the system also provides access to a set of most used tags by other users.
3. The agent chooses a new resource or an existing resource in the system.
4. Agent has two options to choose a tag to be assigned to the selected resource:
 - He plays the role of a speaker and in this case the system plays the role of the interlocutor and we are in a situation of self imitation, ie the system learns words from the agent linguistic knowledge.
 - He plays the role of an interlocutor and in this case the system plays the role of speaker and we are in a situation of social imitation, ie the agent learns from the system by imitating it overall language knowledge stored in the memory of the system.

In both situations, there will be a negotiation between the speaker and the interlocutor to decide the

appropriate tag assignment to the resource in question.

5. It adapts its associative memory by inserting new assignment or modifying previous one by the reinforcement learning.
6. It sends the identifier of the resource and the tag he chose to the system as an input "Ui, Ti, Ri, ti" Where Ui, Ti, Ri, ti represent the user, the tag, the resource, and a time stamp.

B. Results and Discussion

Examples of simulated scenarios are presented by modifying the model parameters (the number of agents, the threshold measure of similarity).

To analyze the emerging language, we use the encoding matrix inspired from [1] and we redefine it as a matrix whose columns are the tags of the system and the lines are system resources, and the intersection of a tag "t" with an resource "r" is the probability to put the resource "r" in the category represented by the tag "t".

Then we use graphic representation to present the encoding matrix content in the form of clusters where each cluster stands for a word (a tag) and the resources that have been tagged by it.

We will present two scenarios of simulation.

1. Example 1

We use a number of agents N = 10 agents, and a threshold of similarity measure S = 0.3 for this example.

We calculate the encoding matrix to analyze the emerging language consisting of emergent categories in collaborative tagging system.

After having calculated the encoding matrix, emergent categories are shown in the form of clusters; the figure below shows some emergent clusters in the tagging system.

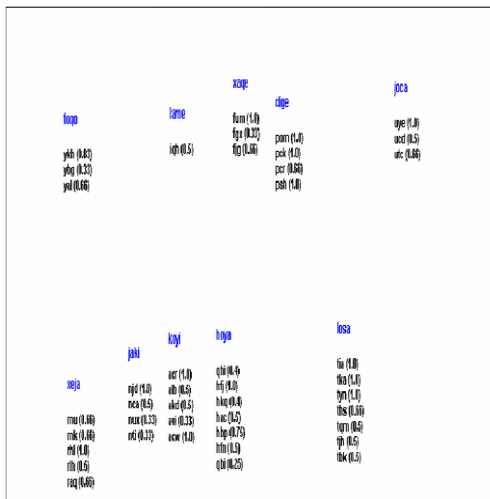


Fig.5 – Some emergent categories

In the encoding matrix, we have an emergent language formed of emergent categories and resources that belong to each category.

In figure 4, in the category represented by the word "xaqe», remember that the tags are randomly generated words, we note that its elements are resources that start with the character "f". For the

agents, these resources belong to a common field. The same thing can be said about the category represented by the tag "joca" and resources starting with "u".

2. Example 2

We use a number of agents N = 100 agents and a threshold similarity measure S = 0.3 for this example. Through linguistic analysis of this example, we calculate the encoding matrix. The following figure shows some emergent clusters in our simulated tagging system.

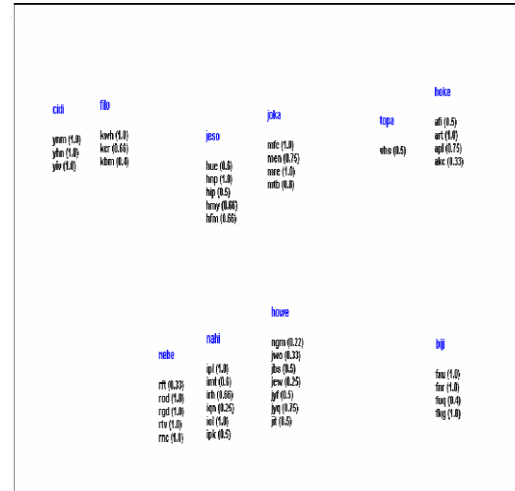


Fig. 6– Some emergent categories

For example, the category represented by the word "jesa" contains resources that start with the character "h".

C. Summary and Discussion

After the results of the simulations mentioned above, a tagging language shared by users of collaborative tagging system, modeled by agents, emerges in the system and in associative memories of the agents since we are interested in the linguistic aspect of this process.

The agents invent words (tags) in order to describe the resources of the system and to put them in different categories.

D. Conclusion

In this work, we studied the dynamics governing the collaborative tagging systems through the first model using a complex systems based modeling and multi agent systems. This modeling approach allows the regeneration of the dynamics observed in the collaborative tagging systems through self-organization and a selections mechanism expressed by the principle of self imitation and social imitation.

Our main objective of this work is the study of the emergence of a language shared by users of tagging system, we treated this objective through the second model, enriching the first model by cognitive structures in the form of associative memories and by using interaction scenarios based on language games, we have shown that users of collaborative tagging system can share a emergent tagging language that has lexical structure.

Tagging language with grammatical structure seems to be a very good perspective and continuation of this work.

The application of these results on existing collaborative tagging systems or the design of new tagging systems taking into account the linguistic aspect of these systems are future goals and a continuation of this work.

The use of other models of the language emergence is another approach to addressing this issue to better understand the cognitive processes of collaborative tagging.

References

- [1] Kaplan, F. L'émergence d'un lexique dans une population d'agents autonomes, Ph.D thesis, Université Paris VI, 2000.
- [2] Golder, S., & Huberman, B. A., The structure of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208, 2006.
- [3] Cattuto, C., Baldassarri, A., Servedio, V. D. P., & Loreto, V., Vocabulary growth in collaborative tagging systems. Arxiv-print: <http://arxiv.org/abs/0704.3316>, 2007.
- [4] Dellschaft, K., Staab, S., An epistemic dynamic model for tagging systems. *Proceedings of the 19th ACM Conference on Hypertext and Hypermedia HT'08* (p. 71-80), 2008.
- [5] Halpin, H., Robu, V., & Shepherd, H., The complex dynamics of collaborative tagging. *Proceedings of the 16th international conference on World Wide Web* (p. 211-220), 2007.
- [6] Santos-Neto, E., Condon, D., Andrade, N., Iamnitchi, A., et Ripeanu, M., "Individual and social behavior in tagging systems," *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, p. 183-192, 2009.
- [7] Steels, L., The synthetic modeling of language origins. *Evolution of Communication* 1.1-35, 1997.
- [8] Steels, L., Kaplan, F., Collective learning and semiotics dynamics, in D. Floreano, J-D. Nicoud, F. Mondada F. (éds.), *Advances in Artificial Life, Lecture Notes in Artificial Intelligence*, p. 704-708, Berlin, Springer- Verlag, 1999.
- [9] Steels, L., *Language Games for Autonomous Robots*. IEEE Intelligent Systems, 2001.
- [10] Steels, L., Kaplan, F., Bootstrapping Grounded Word Semantics, in T. Briscoe (éd.) *Linguistic evolution through language acquisition: Formal and Computational Models*, Cambridge University Press, 2001.

Deployment approaches of wireless sensor networks considering lifetime extension

Soumaya Fellah, Mejd Kaddour and Yahia Lebbah

Laboratoire LITIO, Université d'Oran, BP 1524, El-M'Naouer, 31000 Oran, Algérie,
soumaya.fellah@gmail.com, kaddour.mejdi@univ-oran.dz, ylebbah@gmail.com

Abstract—Sensor placement impacts considerably area coverage and network resources utilization. A key challenge in deployment is to determine a sensor location that optimizes cost by ensuring high sensor coverage. We are interested in a deployment strategy that maximizes the area coverage of the sensors and guarantees at the same time network connectivity between sensors and the sink. In this paper, we propose two deployment approaches that aim to cover a set of deployed targets with the use of minimum number of sensors and ensure network connectivity. The first one relies on integer programming, while the second one is heuristic. Then, we compare these two alternatives in terms of the number of deployed sensors and computation time. Also, we propose an energy efficient approach that ensures the network operation for a specific period of time.

Keywords—*optimization; wireless sensor network; coverage; connectivity; energy consumption; network lifetime; integer programming.*

I. INTRODUCTION

Wireless sensor networks (WSNs) have attracted significant attention due to their integration of wireless, computer, and sensor technology. These networks consist of a multiplicity of nodes that are equipped with processing, communicating and sensing capabilities, and use ad-hoc radio protocols to forward data in a multi-hop mode of operation. The sensors are used to collect, treat and transport data to the sink nodes in the network. Two deployment strategies can be used according to network purpose and environment condition either controlled or random. Controlled deployment is usually applied in accessible environments where the sensors can be placed exactly in the desired location, eventually by using a robot. Random deployment is adapted to inaccessible or hostile environments such as battle fields or disaster regions.

Sensors are self-organized to perform specific tasks such as environment monitoring, target tracking, or infrastructure surveillance. Sensor networks can be used in different applications, such as military applications, environmental applications, health applications, home automation, or commercial applications. Reference [1] gives further detailed about sensor networks applications.

One practical goal of sensor deployment in the design of distributed sensor systems is to achieve an optimal monitoring and surveillance of a targeted region. The optimality of a sensor deployment scheme is a tradeoff between implementation cost and coverage quality levels. An important research problem in wireless sensor networks is the coverage problem, which focuses on how well the field is monitored by sensors [2]. One of the main applications of WSNs is to provide proper coverage of their

deployment regions. Typically, the sensing range is a circular disk centered at the sensor. A given point is said to be covered by a sensor network if it falls in the sensing range of at least one sensor. A region is said to be covered if all the point inside it are covered [3].

In building WSNs, network coverage is one of the fundamental issues. The question is how to deploy the sensor network nodes, under the conditions of ensuring a certain quality of service and achieving the maximization of network coverage with a minimum number of sensor nodes. There are other factors to be considered when a WSN is deployed, such as financial cost.

The target coverage problem is to cover a set of given deployed targets with known coordinates. One of the objectives is to minimize sensing cost. We investigate in the current work coverage problem in sensor networks, under the conditions that both target coverage and network connectivity are satisfied.

As a means, we develop two approaches to determine the sensors positions in order to cover all the targets in the area. The first one is relies on integer programming which provides the optimal sensor placement strategy and the second approach is a heuristic-based algorithm. Both optimal and heuristic approaches aim to minimize the number of sensors placed in area.

Furthermore, this paper investigates the network lifetime problem to ensure the operation of the network for a given predetermined duration. Indeed, a wide dissemination of WSNs is obstructed by the severe energy constraints of individual sensor nodes. This is the reason why a large part of the research in this area focuses on the development of energy consumption efficient strategies. Energy conservation is essential in this context because nodes are usually battery powered.

The remainder of this paper is structured as follows. In Section II, we discuss some previous work in the field of sensor deployment under coverage constraints. In Section III, we present our proposed deployment approaches. In Section IV we describe the lifetime approach. Experimental results are presented in Section V. Finally, we conclude the paper in Section VI.

II. RELATED WORK

Different deployment methods are proposed in order to provide high coverage, [4] developed an algorithm to cope with the sensor placement problem for target location under constraints of the cost limitation and complete coverage, this algorithm is able to find the optimal sensor placement under the minimum cost constraints. Moreover, it can also

find a placement with minimum distance error for large sensor fields. The authors in [5] addressed the issue of covering a set of target points in an area with a finite set of sensors. A probabilistic model is proposed which takes in account the detection probabilities of the sensors which may decay with distance, environmental conditions, and hardware configuration. The objective is to deploy sensors so that the distribution of the sensors meets the probability of detection requirements while minimizing costs. The work in [6] proposed optimized GRG (Sensor self-deployment algorithm Greedy-Rotation-Greedy), for guaranteed coverage radius maximization. [7] Studied optimal deployment in terms of the number of sensors required to achieve four-connectivity and full coverage under different ratios of sensors' communication range to their sensing range. It proposed a new deployment pattern called the Diamond pattern, which can achieve four-connectivity and full coverage. [8] Presented a novel algorithm that employs grouping and time scheduling to turn off some of the sensors covering of them on the entire area. In this algorithm the sensing range of the nodes depends on their energy. In [9], authors investigated random and deterministic node deployments for large-scale wireless sensor network under different performance metrics such as coverage and energy consumption. They considered three alternatives: a uniform random, a square grid, and a pattern-based Tri-Hexagon Tiling node deployment and calculate exactly the k-covered points. [10] Proposed solutions to the minimum cost coverage problem under random and deterministic deployments. The aim of this work is to determine the minimum number of sensors required to cover the desired arbitrarily shaped geometric profile and not to propose the optimal placement strategy with an arbitrary amount of sensors. In [11], authors considered a probabilistic sensing model that provides different sensing capabilities in terms of coverage range and detection quality with different costs. A sensor deployment problem for a planar grid region is formulated as a combinatorial optimization problem with the objective of maximizing the overall detection probability within a given deployment cost. In [12], deployment strategies for two-dimensional and three-dimensional communication architectures for underwater acoustic sensor networks are proposed, and a mathematical deployment analysis for both architectures is provided. The objective is to determine the minimum number of sensors to be deployed to achieve optimal sensing and communication coverage. Finally the aim of [13] is to minimize the number of sensors deployed, while providing satisfactory quality of data and provide enough coverage for every point on the grid at the same time.

III. PROPOSED DEPLOYMENT APPROACHES

Consider a region to be monitored using a sensor network. The sensors are deployed over the region in a deterministic fashion so that each target can be monitored by at least one sensor.

In order to collect information from the sensor nodes to monitoring center, we assume that there is a communication path towards the sink for each deployed sensor,

In this section, we focus on the design of two approaches that can provide full target coverage using a minimum number of sensors. In the first one, we propose a linear

integer program which provides an optimal placement strategy of sensors and the connection between all deployed sensors and the sink. In the second approach, we propose a heuristic algorithm.

A. Proposed linear approach

Given a set of n targets, denoted as $T = \{t_1, t_2, \dots, t_n\}$ with predetermined positions, the objective is to find a set of sensor positions that cover them. We consider a square grid as a deployment region. We define below the main parameters:

- c : number of columns in the grid
- l : number of lines in the grid.
- r_s : sensing range of sensors.
- r_c : radio communication range of sensors.
- y_{ij} : is a binary denoting if a sensor deployed on coordinates (i, j) is connected to the sink.
- l_0 : sink location.

The decision variables of our optimization model are defined as follows:

- x_{ij} : is a binary denoting if there is a sensor is deployed on coordinates (i, j) of the grid.
- α : number of deployed sensors.

The parameter $d(l_a, l_b)$ denotes the Euclidean distance between two possible locations l_a and l_b . Let $\sigma(x_{ij}, t)$ be an indicator function of whether a target t can be covered by a sensor located on the location (i, j) , it can be defined as:

$$\sigma(x_{ij}, t) = \begin{cases} 1, & \text{if } (d(l_{ij}, l_t) \leq r_s) \text{ where } l_{ij} \text{ is location of site } (i, j) \\ & \text{and } l_t \text{ is location of target } t. \\ 0, & \text{otherwise} \end{cases}$$

Besides, let $\delta(l_a, l_b)$ denotes an indicator function of whether a sensor deployed on location a can establish a direct communication link with a sensor deployed on location b .

$$\delta(l_a, l_b) = \begin{cases} 1, & \text{if } (d(l_a, l_b) \leq r_c) \text{ where } l_a \text{ is location of sensor } a \\ & \text{and } l_b \text{ is location of sensor } b. \\ 0, & \text{otherwise} \end{cases}$$

Now, our optimization model can be stated as follows:

Minimize α

$$\text{Subject to } \sum_{i=0}^{l-1} \sum_{j=0}^{c-1} \sigma(x_{ij}, t) \geq 1 \quad \forall t \in T \quad (1)$$

$$y_{ij} \leq y_{i'j'} \quad \forall (i, j), \forall (i', j') \text{ with } d(l_{ij}, l_{i'j'}) \leq r_c \text{ and } d(l_{ij}, l_0) > r_c \quad (2)$$

$$y_{ij} = x_{ij} \quad \forall (i, j) \quad (3)$$

$$\sum_{i=0}^{l-1} \sum_{j=0}^{c-1} x_{ij} = \alpha \quad (4)$$

The objective function minimizes the number of deployed sensors. The constraint (1) requires that all targets are covered by at least one sensor. The constraint (2) requires that the Euclidean distance between deployed sensor (not directly connected to the sink) and an other already connected sensor is lower or equal to r_c . The constraint (3) ensures that all deployed sensors are connected. The constraint (4) calculates the number of deployed sensors.

B. Proposed heuristic approach

In the first time, we select one target to be covered and in second part, connection path is built with the aim to reach sink node. It relies on an iterative process, where at each iteration we select a sensor location that covers the maximum number of yet uncovered targets, while ensuring that the distance between this location and the sink or one of already deployed sensors is less or equal to the communication range. This process is repeated until all targets are covered.

The approach steps are described in the following algorithm:

Algorithm 1. Heuristic Approach

Input:

$T = \{t_0, \dots, t_n\}$ \triangleright set of uncovered targets

$C = \{\text{sink}\}$ \triangleright connected sensors

Output: number of deployed sensors and their positions

While $T \neq \emptyset$ do

1. Select $t \in T$ where t has a minimum Euclidean distance with a connected sensor.

2. Select a sensor position s that covers a maximum number of targets where Euclidean distance with a connected sensor is less than communication range. This position has a minimum distance with one of connected sensor.

3. $T = T - \{t\}$ T' : \triangleright set of targets covered by sensor s .

4. $C = C \cup \{s\}$

5. If t is covered goto 1 else goto 2.

End while

IV. LIFETIME APPROACH

In this section, a novel energy efficient approach is proposed. The main difference with other approaches is that this one ensures a network operation for a specific lifetime. This is performed by reinforcement of sensors disable to pursue his tasks for apredetermined period.

In Subsection A, the energy model adopted is presented, while in Subsection B, lifetime approach is described.

A. Energy model

In this work, the energy model adopted is defined by the following equations according to [14]:

$$E_{Tx}(k,d) = T_{Tx-elec}(k) + E_{Tx-amp}(k,d) = E_{elec} \cdot k + E_{amp} \cdot k \cdot d^2 \quad (6)$$

$$E_{Rx}(k) = E_{Rx-elec}(k) = E_{elec} \cdot k \quad (7)$$

Where $E_{Tx}(k,d)$ denotes the amount of energy required to transmit a message of k bits over a distance d .

Similarly, the energy dissipated by a sensor for the reception $E_{Rx}(k)$ of a message of k bits.

We have set $E_{elec} = 50\text{nJ/bit}$ and $E_{amp} = 100\text{pJ/bit/m}^2$

B. Lifetime approach description

Energy consumption is a primary concern in wireless sensor networks. This is because in many practical scenarios, sensor node batteries cannot be easily refilled, and nodes have a finite lifetime.

The aim of many sensors applications is to ensure target coverage for a given duration even with degradation of sensing rate that we note α .

This paper presents an approach that provides energy efficiency to extend the network lifetime to a specific period.

Sensors node cannot ensure a given lifetime are divided in two classes. The first one is a set of sensor able to ensure a given lifetime with a degradation of sensing rate. The second class is a set of sensors cannot ensure a given lifetime even with a degradation of sensing rate, this class represents sensors that need a important energy capacity to ensure a given lifetime.

Energy expenditure of the sensor nodes occurs during the wireless communication, in order to reduce the transmissions number, we propose to use a upper sensing rate for the first class that reduces the total number of data transmission which decreases energy consumption.

Sensor of the second class must be strengthening with redundant sensor. The last one pursues the task of shutdown sensor.

V. NUMERICAL RESULTS

In this subsection we present some numerical results of our approaches. In the following present the performance results and comparison between optimal and heuristic approaches in terms of number of deployed sensors and run time.

We have calculated the number of deployed sensors and run time with several topologies. The simulated environment consists of 5, 10, 15, 20 targets distributed over a 10×10 grid. This number is calculated for 5 scenarios.

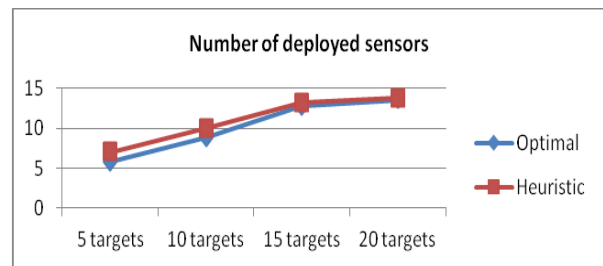


Figure 1. The number of deployed sensors

Figure presents the number of deployed sensors under different number of targets. The number of deployed sensor enhance with the increase of the number of targets.

Results show that the targets distribution has an effect on the number of deployed sensors.

Although linear provides the optimal solution, we observe that no significant difference between optimal and heuristic approach.

In this subsection we compare a run time to have a solution in the both approaches.

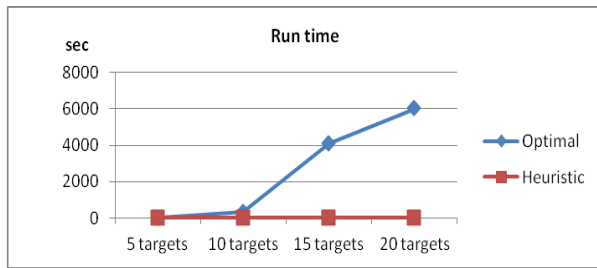


Figure 2. The run time

Previous figure illustrates a large difference in the run time between the approaches. Heuristic approach takes a very short time to provide a deployment strategy. Optimal approach searches all possible solutions and selects the best one what generates a large run time.

To evaluate the performance of the proposed scheme, we have compared proposed lifetime approach with the approach without strengthening of large energy consumption sensor.

Figure 3 presents network lifetime without strengthening of sensor. The result obtained by measuring the time until the first node shutdown. Our proposed approach improve network lifetime from 67% to 100%.

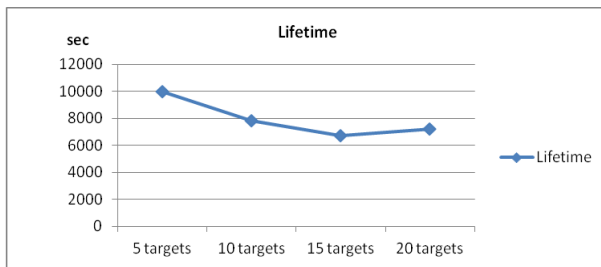


Figure 3. Lifetime

Figure 4 illustrates a number of shutdown sensors, it represents a set of sensor cannot ensure a given lifetime.

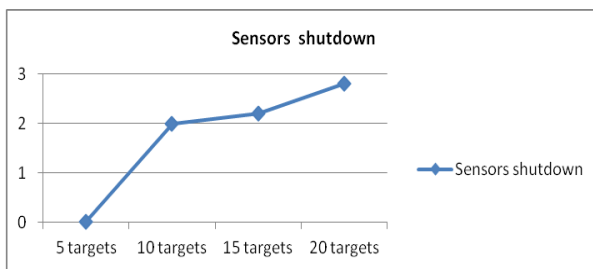


Figure 4. Number of sensors shutdown

Following figures show how additional sensors affect the lifetime under different values of α . Proposed approach can improves the lifetime from 52% to 100% by adding one redundant sensor.

A near superlative rate affects the number of additional sensors that reduces with α increase.

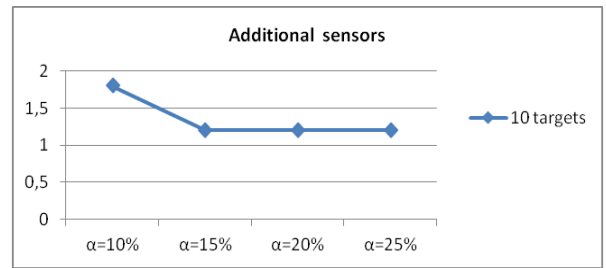


Figure 5. Number of additional sensors for 10 targets

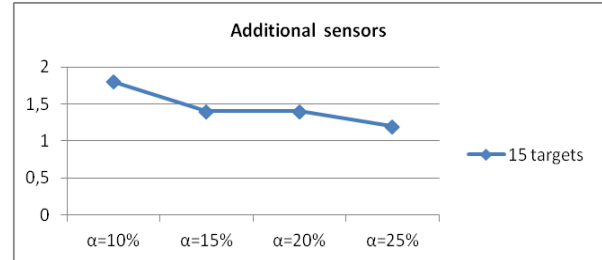


Figure 6. Number of additional sensors for 15 targets

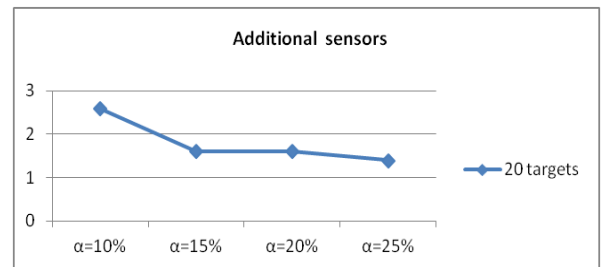


Figure 7. Number of additional sensors for 20 target

VI. CONCLUSION

The present work investigates the coverage problem in the wireless sensor network. We focus on the target coverage and the network connectivity. We have proposed ainteger and heuristic approaches with the aim of find a minimum number of sensors such that each target is covered at least a sensor and the selected sensors is connected to the sink.

Our approaches minimize the number of sensors witch decrease the network cost.

The results show that the targets positions, the number of targets have an important effect on the number to deployed sensors that cover all targets. Comparison of both optimal and heuristic approach shows a not important difference in term of number of deployed sensors.

Heuristic approach provides a near optimal deployment strategy in very short time this privilege is very important in many sensor network applications.

Proposed lifetime approach ensures a given lifetime with the use of additional sensors and a sensing rate degradation that reduces the total number of data transmission.

REFERENCES

- [6] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci.: "Wireless sensor networks: a survey", Computer Networks 38 (2002) 393–422.
- [7] Ahmed, Nadeem and Kanhere, Salil S. Jha Sanjay.: "A Pragmatic Approach to Area Coverage in Hybrid Wireless Sensor Networks", Wireless Communications and Mobile Computing 23-45, January 2011.

- [8] Wei Wang, VikramSrinivasan, Kee-Chaing Chua.: "Coverage in Hybrid Mobile Sensor Networks" ,IEEE transactions on mobile computing, vol. 7, no. 11, november 2008.
- [9] Frank Y. S. Lin and P. L. Chiu.: "A Near-Optimal Sensor Placement Algorithm to Achieve Complete Coverage/Discrimination in Sensor Networks", IEEE COMMUNICATIONS LETTERS, VOL. 9, NO. 1, JANUARY 2005.
- [10] Brian Carter, RammohanRagade.: A Probabilistic Model for the Deployment of Sensors, 2008.
- [11] Xu Li, Hannes Frey, Nicola Santoro, Ivan Stojmenovic.: "Localized Sensor Self-deployment for Guaranteed Coverage Radius Maximization", inproceedings of the 2009 IEEE international conference on Communications. ISBN: 978-1-4244-3434-3, 2009.
- [12] XiaoleBai, Ziqiu Yun, Dong Xuan, Temn H. Lai, WeijiaJia.: "Optimal Patterns for Four-Connectivity and Full Coverage in Wireless Sensor Networks", IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 9, NO. 3, MARCH 2010.
- [13] J. Shanbehzadeh, M. Mehrani, A. Sarrafzadeh, Z. Razaghi.: "An Energy Efficient Coverage Method for Clustered Wireless Sensor Networks", proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol II,IMECS 2010, March 17-19, 2010, Hong Kong. ISBN:978-988-18210-4-1, ISSN: 2078-0958.
- [14] Wint Yi Poe, Jens B. Schmitt.: "Node Deployment in Large Wireless Sensor Networks: Coverage, Energy Consumption, and Worst-Case Delay".AINTEC'09, November 18–20, 2009, Bangkok, Thailand.
- [15] S. A. R. Zaidi, S. A. Khayam, D. C. McLernon.: "On Minimum Cost Coverage in Wireless Sensor Networks" , 43rd Annual Conference on Information Sciences and Systems (CISS 2009), John Hopkins, NJ, U.S.A, March 2009.
- [16] Qishi Wu, Nageswara S.V. Rao, Xiaojiang Du, S. Sitharamalyengar, Vijay K. Vaishnavi.: "On efficient deployment of sensors on planar grid", Computer Communications 30 (2007) 2721-2734.
- [17] Dario Pompili, TommasoMelodia, Ian F. Akyildiz.: "Three-dimensional and two-dimensional deployment analysis for underwater acoustic sensor networks", Ad Hoc Networks 778-790, 2009.
- [18] ZhengRuan.: "Wireless Sensor Network Deployment in Mobile Phones Environment", thesis November 2009.
- [19] Stefanos A. Nikolidakis, Dionisis Kandris, Dimitrios D. Vergadosand Christos Douligeris.: "Energy Efficient Routing in Wireless Sensor Networks ThroughBalanced Clustering", ISSN: 1999-4893, 2013.

Proposition D'un Protocole De Communication Dans Un Réseau Hybride Pour Améliorer La Sécurité Routière

T.BENSIRADJ, S. MOUSSAOUI, S.HASNI

Département d'informatique, Université des Sciences et de la Technologie Houari Boumediene USTHB, Code postal 32

El Alia, 16111 Bab Ezzouar, Algérie

ayabn7@gmail.com

moussaoui_samira@yahoo.fr

Résumé— Parmi les principaux problèmes des protocoles de communication entre un WSN et un VANET dans les HSVNs, est celui du délai de livraison des messages. Ce problème est dû à plusieurs causes: les courtes portées de transmission et de détection du capteur, la forte mobilité des véhicules, la présence d'obstacles, l'interférence, les capacités limitées du stockage et du traitement du capteur et la faible quantité d'énergie dans les WSNs ...etc. En plus il y a deux types de messages utilisés dans ce type de protocole de communication qui sont: les messages utiles échangés entre les deux réseaux (informations routières: comme la densité des véhicules, les conditions météorologiques, la présence d'obstacles). Et les messages critiques envoyés par les capteurs aux véhicules (informations d'avertissements utilisées pour éviter une situation dangereuse). Ces derniers exigent un délai de transmission optimal. Pour réaliser un protocole dans cette catégorie des protocoles de communication, il est indispensable de prendre en considération tous ces problèmes et les types des messages. Notre travail est une proposition de protocole de communication entre un WSN et un VANET nommé MEP (Message Exchange Protocol). Ce dernier assure: un échange d'informations routières (messages utiles) entre les deux réseaux, dans le temps disponible où le véhicule est dans la portée de transmission du capteur et une livraison urgente des messages critiques par les capteurs aux véhicules.

Mots- Clés— WSNs, VANETs, HSVN.

I. INTRODUCTION

Les réseaux de capteurs sans fil (WSNs) ont prouvé leur contribution dans plusieurs domaines (militaire, médical, environnemental, industriel...etc.). Parmi les domaines où le rôle d'un WSN est très important, les systèmes de transport intelligent (STI). Ces derniers sont déployés pour plusieurs objectifs: la gestion et le contrôle du trafic routier, le paiement électronique, l'organisation des voyages publics, le contrôle et le respect de la réglementation et la sécurité routière. Cette dernière peut être améliorée en utilisant un WSN comme un outil (présence d'obstacles, animaux, conditions routières difficiles...etc.) qui aide les conducteurs à la diminution du nombre d'accidents. Ce résultat peut être efficace si un WSN collabore avec un autre type de réseau, comme un réseau véhiculaire (VANET). Une nouvelle génération de réseau HSVN [1,2,4,5,6] (Hybrid Sensor and Vehicular Networks) a été créée en combinant les deux réseaux travaillant conjointement dans un cadre permettant la diminution Maximale du nombre d'accidents.

Les protocoles de communication dans un HSVN peuvent être répartis en trois classes: les protocoles de communication à l'intérieur d'un WSN, les protocoles de communication à l'intérieur d'un VANET et les protocoles de communication entre un WSN et un VANET.

Un protocole de communication entre un WSN et un VANET dans un HSVN est caractérisé par le modèle suivant: un véhicule traverse différentes routes pour atteindre sa destination. Il collabore avec les WSNs déployés sur les bords des routes afin

d'obtenir des informations sur les différentes routes dans la direction de sa destination.

Quand un véhicule se présente dans la portée de transmission d'un capteur passerelle (il joue le rôle d'un relais entre le WSN et le VANET). Ce dernier envoie des informations (collectées par les capteurs et les véhicules et stockées dans sa base de données) sur l'état des routes suivantes dans la direction de déplacement du véhicule s'orientant vers sa destination (le capteur détermine la destination du véhicule à travers un type de message envoyé par le véhicule contenant son identificateur et les coordonnées de sa destination). À la réception de ces informations par le véhicule, ce dernier transmet à son tour au capteur les informations collectées grâce à d'autres capteurs ou d'autres véhicules. Ce type d'informations est transmis par des messages particuliers appelés messages utiles portant des informations non critiques mais importantes pour éviter les causes (état météorologique des routes, densité des véhicules, présence d'obstacles...) qui pourraient mener à des accidents, des situations dangereuses ou de mauvaises conditions de circulation du trafic routier (éviter la circulation et l'encombrement, organisation des transports publics...etc.).

D'autre part, il existe un autre type d'informations détectées par les WSNs sur les routes. Ces informations représentent des messages d'alerte, c'est-à-dire que les véhicules doivent être informés, le plus vite possible d'une situation dangereuse afin d'éviter toute complication de cette situation. Ce type d'informations transmet à travers des messages que nous désignons par messages critiques. Pour ces derniers, le délai de transmission est un paramètre critique à considérer car ils portent des informations urgentes dépendantes des délais. (Par exemple, si un animal traverse la route quand un véhicule se dirige vers sa position, il faut qu'il soit informé suffisamment à temps pour le freinage du véhicule).

Dans le cas d'échange des messages utiles, le temps disponible (où le véhicule est dans la portée de transmission d'un capteur) pour faire l'échange d'informations entre le capteur et le véhicule, dépend de la durée du traitement des messages portant une quantité importante d'informations. Par contre dans le cas des messages critiques, la durée du traitement est négligeable par rapport au temps disponible pour transmettre ces derniers aux véhicules. Cependant la stratégie suivie par le protocole d'échange joue un rôle important pour réduire le délai d'arrivée des messages critiques aux véhicules.

Cet article est organisé comme suit: dans la section II, nous allons présenter les approches utilisées pour réaliser les protocoles de communication entre un WSN et un VANET dans un HSVN et les concepts de base utilisés par notre protocole. Ensuite dans la section III, nous allons expliquer le fonctionnement de notre protocole. Dans la section IV, nous présentons les résultats de simulation. Nous terminons cet article par une conclusion dans la section V.

II. APERÇU DES TRAVAUX CONNEXES ET CONCEPTS DE BASE DU PROTOCOLE MEP

1. Aperçu des travaux connexes

Il existe deux approches pour réaliser ces protocoles de communication. La première approche [1,2,3] se base sur

une communication à deux phases. La première phase est utilisée pour initialiser la connexion entre le capteur et le véhicule. Cette dernière commence quand le capteur détecte le véhicule. La deuxième phase est utilisée pour l'échange des messages (utiles, critiques). La deuxième approche [4,5] utilise les messages beacon diffusés périodiquement par les véhicules. Ces derniers sont utilisés pour la découverte du voisinage dans les réseaux véhiculaires en plus de ce rôle, on leur a ajouté une autre fonction. Cette fonction est représentée par l'intégration d'autres informations (vitesse, identificateur, coordonnées de destination...) du véhicule dans le message beacon. Cette approche utilise ce type de message car la portée de transmission du véhicule est plus grande que celle du capteur ce qui permet d'éviter l'étape d'initialisation de la connexion entre les deux. Car le capteur peut commencer le travail de la collaboration des qu'il reçoit le message beacon.

Concernant les protocoles de communication à deux phases, les points communs entre les propositions sont: la 1ere phase est utilisée pour initialiser la connexion entre les deux. Dans cette dernière, le capteur passerelle envoie une demande de connexion au véhicule. Ce dernier répond par un accusé de réception contenant son identificateur et les coordonnées de sa destination. La 2eme phase est utilisée pour l'échange des messages, en premier le capteur passerelle envoie des messages au véhicule. Après ce dernier lui envoie à son tour des messages. Le temps disponible pour faire l'échange est donné par l'équation 1. Les auteurs ont travaillé sur l'augmentation de ce temps tout en respectant les contraintes de ces protocoles.

- $T_{dispo} = 2 * R_c / V - (T_{d\acute{e}tecte} + T_{traitement})$ secondes (1).

R_c : Laportée de transmission du capteur

V : La vitesse du véhicule.

$T_{d\acute{e}tecte}$: Le temps de détection du véhicule.

$T_{traitement}$: Le temps de traitement et de préparation des messages.

Les auteurs ont travaillé sur la réduction de la quantité d'informations échangées entre les deux réseaux. En plus ils ont pensé à un codage simple de données pour réduire le temps de traitement de l'information ($T_{traitement}$). Les avantages des travaux de recherche peuvent être résumés dans les points suivants: Dans la proposition [1], un modèle de segmentation des routes a été proposé. Ce dernier permet de décomposer une route en un ensemble de segments numérotés pour faciliter la localisation des véhicules et des événements routiers. De plus dans la proposition [2], il ajoute à ce concept un modèle de données permettant le codage d'informations routières sur un segment de route en quelque nombre d'octets. Et une approche groupée est utilisée par les deux réseaux pour l'économede la consommation d'énergie des capteurs et l'élimination des traitements inutiles. La segmentation des routes et le modèle de données permettent aussi la réduction de la quantité d'informations stockées dans les bases de données du capteur. D'autres auteurs [3] ont pensé à l'utilisation d'une approche distribuée dans les réseaux de capteurs. Cette approche permet la rédaction de la quantité des informations stockées dans un capteur. De plus elle évite les traitements complexes.

Cependant, l'approche à deux phases a un inconvénient majeur sur le délai d'échange des messages entre les deux réseaux. Ce délai représente la métrique la plus importante

dans ce type des protocoles. L'utilisation de la première phase prend un temps supplémentaire de temps disponible de l'échange Eq(2).

- $T_{dispo} = (1) - T_{init}$ secondes (2).

T_{init} : le temps d'initialisation de la connexion entre le capteur et le véhicule.

De plus à ce point négatif de cette approche, nous pouvons donner des inconvénients liés aux solutions présentées dans les paragraphes précédents. Ces inconvénients sont:

1. Dans toutes les solutions [1,2,3], il n'y a pas une stratégie permettant la distinction entre un message utile et un message critique (l'envoi des messages se fait par rapport à l'ordre d'arrivée). Malgré que les messages critiques sont prioritaires que les messages utiles.
2. Dans la solution [1], la communication entre le capteur passerelle et tous les véhicules cause deux problèmes: la consommation élevée d'énergie du capteur et la duplication des messages surtout dans le cas d'une forte densité des véhicules.
3. La solution [2], dans le cas où l'échange des messages entre le capteur passerelle et le véhicule (chef de groupe) ne se termine pas avec succès. Le capteur passerelle doit attendre un nouveau chef de groupe véhiculaire, ce qui entraîne à un problème de délai pour la livraison des messages surtout dans le cas d'une faible densité des véhicules.
4. Dans [3], comme les informations sont réparties sur plusieurs capteurs passerelles, le temps de réponse du capteur sur une demande d'un véhicule, sera long (surtout dans le cas d'un problème de communication à l'intérieur d'un WSN). De plus la mise à jour des bases de données distribuées est difficile.

Pour éviter l'inconvénient de l'approche à deux phases, la deuxième approche a été proposée. Elle utilise le message beacon comme un porteur des informations du véhicule. Quand le capteur reçoit ce message, il prépare les messages qui seront envoyés au véhicule. À la détection du véhicule par ce dernier, il lui envoie les messages. Donc on a évité la phase d'initialisation de la connexion. Les auteurs ont utilisé les messages beacon, pour les deux raisons suivantes: 1ere la portée de transmission du véhicule est plus grande que celle du capteur, alors théoriquement le message beacon est reçu par le capteur avant que le véhicule ne soit détecté par ce dernier. Et les messages beacon sont diffusés périodiquement dans un intervalle court, ce qui augmente la probabilité de leur réception par les capteurs.

Dans cette approche, les auteurs ont travaillé sur la réduction du temps de traitement et de préparation des messages par les capteurs. Par exemple dans la solution [5], les messages beacon sont utilisés comme des porteurs des informations du véhicule et des informations utiles envoyées au capteur. L'avantage de cette proposition est quand le capteur détecte le véhicule l'échange d'information se fera dans un seul sens (capteur vers véhicule). Car le capteur a déjà reçu les informations apportées par le véhicule. Pour éviter la surcharge (informations routières envoyées par le véhicule au capteur) ajoutée au message beacon, d'autres auteurs ont pensé à ajouter quelques informations à ce dernier. Ces informations sont par exemple l'identificateur du véhicule, les coordonnées de sa destination et sa vitesse actuelle. On trouve ce concept dans la solution [6]. Cette dernière se base sur la notion de groupe (les deux réseaux sont respectivement répartis en groupes). Elle propose un modèle de message permettant la distinction entre un message utile et un message critique. L'avantage de cette solution est ce modèle de message et

l'utilisation d'autres véhicules pour récupérer les messages critiques.

Cependant, l'approche utilisant les messages beacon a un inconvénient majeur sur le délai d'échange des messages entre les deux réseaux dans le cas d'une forte collision ou d'une interférence. C'est-à-dire si le message beacon diffusé par le véhicule, n'est pas reçu par le capteur ou il fait un retard important pour arriver à ce dernier, les protocoles utilisant cette approche pourront avoir des problèmes dans leur fonctionnement.

De plus à ce point négatif de cette approche, nous pouvons donner des inconvénients liés aux solutions présentées dans les paragraphes précédents. Ces inconvénients sont:

1. Dans le cas où le message beacon n'est pas reçu par le capteur ou il arrive après la détection du véhicule, il n'y a pas un moyen utilisé par les solutions [5,6] permettant de rattraper le temps perdu.
2. Dans la solution [5], la communication entre le capteur passerelle et tous les véhicules cause deux problèmes: la consommation élevée d'énergie du capteur et la duplication des messages surtout dans le cas d'une forte densité des véhicules. De plus la quantité d'informations ajoutées aux messages beacon, peut causer un problème d'inondation dans le réseau véhiculaire (car ces messages sont diffusés périodiquement par tous les véhicules du réseau véhiculaire).
3. De même dans la solution [5], il n'y a pas une stratégie permettant de distinguer un message utile d'un message critique (l'envoi des messages se fait par rapport à l'ordre d'arrivée). Malgré que les messages critiques sont prioritaires que les messages utiles.
4. Dans la solution [6], l'utilisation des capteurs passerelles (chefs des groupes des réseaux de capteurs) uniquement pour transmettre les messages critiques aux véhicules, peut causer un problème dans le délai d'arrivée de ces messages. Par exemple: si un événement critique se produit dans un segment de route proche de la position du déplacement du véhicule et il n'y pas un capteur passerelle, le véhicule sera informé seulement après qu'il traverse ce segment.
5. De même dans la solution [6], il n'y a pas un moyen permettant de trier les messages critiques selon leur degré d'importance (les messages critiques sont envoyés au véhicule selon leur ordre d'arrivée au capteur passerelle).

Dans la sous section précédente, Nous avons présenté les approches utilisées pour réaliser un protocole de communication entre un réseau de capteurs sans fil et un réseau véhiculaire dans un réseau hybride (HSVN). Chaque approche a des avantages et des inconvénients. Et chaque solution proposée dans ces derniers vise à assurer l'échanges des messages routiers entre les deux réseaux dans le temps disponible. Notre étude nous a permis de déduire la métrique la plus importante dans ce type de protocole qui est le délai d'échange. Comme il existe d'autres métriques dans ce dernier (la consommation d'énergie des capteurs...). Nous avons basé notre travail sur le délai d'échange par une proposition d'un protocole de communication entre un WSN et un VANET nommé MEP (Message Exchange Protocol) qui assure un délai plus rentable par rapport à ceux obtenus dans les solutions existantes. Ce protocole a pour objectif: l'échange des messages utiles dans le temps disponible et la livraison des messages critiques par les capteurs aux véhicules dans un

temps optimal inférieur au temps disponible. Dans la sous section suivante, nous allons donner les concepts de base utilisés par notre protocole.

2. Concepts de base du protocole MEP

2.1. Approche utilisée par le protocole

Nous avons utilisé la notion de groupe pour les deux réseaux. Pour les WSNs, une topologie linéaire groupée et des véhicules répartis en groupes en utilisant des algorithmes de groupage connus dans le domaine des réseaux véhiculaires. Notre protocole est un protocole hybride des deux concepts des deux approches présentées dans la sous section II.1.

2.2. Informations routières échangées

Nous avons supposé qu'une région géographique est décomposée en un ensemble de zones. Chaque zone contient un nombre donné de segment. Le nombre et la position des capteurs sont liés à la nature de la zone (de faible danger, de moyen danger ou de fort danger). Pour chaque segment, nous avons proposé un modèle simple de données (utilisé par les capteurs et les véhicules) permettant le codage d'informations routières sur ce dernier en quelque nombre d'octets. Par exemple deux bits pour les informations suivantes (état climatique, densité des véhicules et présence d'obstacles).

La quantité d'informations d'une zone, contenant un nombre donné de segments, est donnée par l'équation (3).

$$\bullet \quad Qz = (x + 6 + \text{taille}(D)) * 2^x + y \text{ bits}(3).$$

2^x : Le nombre des segments x : La taille d'un segment

y : La taille d'une zone

taille(D) : La taille de la date d'enregistrement de l'événement. Ce champ est très important car si la date de l'événement est ancienne, le capteur élimine cet événement (car la capacité de stockage dans le capteur est limitée).

Ce modèle de données permet de coder une quantité importante d'informations routières dans un espace de stockage limité, ce qui soulage le capteur d'une base de données chargées et des traitements complexes. Donc il évite la consommation élevée d'énergie.

2.3. Mode de fonctionnement des capteurs et protocoles MAC

Chaque capteur peut fonctionner en mode de détection, de routage ou de passerelle (la figure 1). Il est équipé par une interface 802.15.4 pour communiquer avec les capteurs et les véhicules. Ces derniers sont équipés par la même interface et une interface 802.11p pour qu'ils communiquent entre eux.

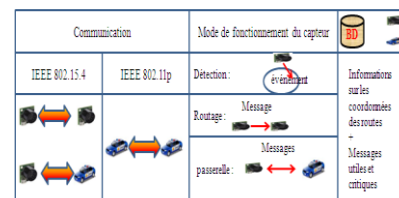


Figure 1: Mode de fonctionnement des capteurs et les protocoles MAC

2.4. Types des messages échangés entre les deux réseaux

-MB: représente le message beacon diffusé périodiquement par les véhicules.

Format général du Message beacon: Message beacon + les informations suivantes:

-Coordonnées de la destination du véhicule.

-Identificateur du véhicule pour la sécurité.

-Bit CH (1: Chef de groupe, 0 : Non chef de groupe).

-Bit indicateur (0: pour informer le capteur qui a reçu ce message beacon que le véhicule ne lui envoie pas un message utile. 1:

informer le capteur qui a reçu ce message beacon que le véhicule va lui envoyer un message utile). Le véhicule positionne ce bit à 1 quand il entre dans une nouvelle zone (le véhicule connaît qu'il est dans une nouvelle zone à l'aide de changement obtenu par son dispositif de localisation).

-MU: représente le message utile envoyé par le véhicule CH au capteur passerelle (nous avons exigé que le véhicule envoie les informations utiles aux capteurs passerelles dans un message utile MU sans qu'elles soient ajoutées aux messages beacon pour éviter leur surcharge (problème d'inondation dans un VANET). De plus les informations utiles sont uniquement envoyées par les chefs des groupes véhiculaires ou les véhicules relais aux capteurs passerelles.

-MSG: représente le message utile envoyé par le capteur passerelle au véhicule CH. Ce message peut être envoyé au véhicule Non CH par le capteur passerelle dans le cas où il est toujours enregistré dans sa base de données et à cet instant un véhicule non CH est détecté par ce dernier.

-MC: représente le message critique envoyé par un capteur à un véhicule.

-ACK : envoyé par le véhicule CH ou véhicule Non-CH, quand ils reçoivent les messages (MSG, MC) envoyés par le capteur.

III. PROTOCOLE D'ÉCHANGE DES MESSAGES (MEP)

1. Les messages utiles

Dans l'échange des messages utiles, Nous avons déduit une de ces trois situations suivantes:

- Échange des messages utiles se produit uniquement entre le capteur passerelle et le véhicule CH (pour économiser la consommation d'énergie des capteurs passerelles) en utilisant un protocole basé sur les messages beacon (l'échange se termine avec succès) (Cas A).
- Le capteur passerelle envoie les messages utiles au véhicule (relais) qui lui a envoyé en premier son message beacon. Après que ce capteur a reçu le message beacon du véhicule CH, dans le cas où: l'échange entre le capteur et le véhicule (CH) ne se termine pas avec succès (Cas B).
- Le capteur passerelle échange les informations avec le véhicule à base d'une communication à deux phases. S'il y a un problème d'interférences ou de collision (le message beacon n'est pas arrivé au capteur passerelle ou il arrive après la détection du véhicule par ce dernier) (Cas C).

Pour chaque situation mentionnée au-dessus, nous allons expliquer comment le protocole MEP fonctionne dans les sous sections suivantes.

1.1 Cas A: échange entre le capteur passerelle et le véhicule (CH)

1.1.1 Véhicule CH entre dans une nouvelle zone z (Figure 2)

Véhicule (CH): Dans ce cas, il prépare un message MU contenant les informations des segments de routes des zones précédentes. Ensuite il diffuse un message beacon contenant les coordonnées de sa destination, son identificateur, le bit CH (dans ce cas bit CH = 1) et le bit indicateur (dans ce cas égal à 1). Après il envoie le message MU au capteur passerelle.

Capteur passerelle : il reçoit ce message quand il est dans la portée de transmission du véhicule (CH). Il traite le message beacon. En premier, il teste le bit CH (dans ce cas bit CH = 1) et le bit indicateur (dans ce cas le bit indicateur = 1). Alors, il se met à l'attente de l'arrivée d'un message MU contenant les différentes informations des segments de route des zones précédentes. Pour chaque segment, on a deux cas, soit il y a des

informations qui ne sont disponibles que dans la base de données du véhicule, soit l'heure d'enregistrement de l'un des états de ce segment dans le véhicule est plus récente que celle de la base de données du capteur. Dans ces deux cas, le capteur passerelle met à jour sa base de données. Ensuite le capteur passerelle insert les informations des zones suivantes dans la direction du véhicule dans le message MSG à la base des coordonnées de la destination du véhicule par rapport aux informations disponibles dans sa base de données.

Le temps de traitement : du message beacon et du message MU, (respectivement diffusé et envoyé par le véhicule au capteur passerelle) et la préparation du message MSG par ce capteur, est (Traitement). La valeur de ce temps, est liée aux deux cas suivants:

1^{er} cas : le traitement des messages se termine avant que le véhicule ne soit détecté par le capteur passerelle dans ce cas:

Ttraitement = Tdétecte – Tarrivé (Ttraitement = 0).

Tarrivé: le temps d'arrivée du message beacon + le temps d'arrivée du message MU au capteur.

2^{eme} cas : le traitement des messages ne se termine pas, après que le véhicule soit détecté par le capteur dans ce cas:

Ttraitementr = Ttraitement - (Tdétecte – Tarrivé).

Ttraitementr : le temps restant pour terminer le traitement des messages.

Comparaison entre les temps des traitements dans les deux approches

Nous remarquons que dans les deux cas, le temps disponible dans un protocole utilisant un message beacon est toujours dans la majorité des cas supérieur au temps disponible d'un protocole à deux phases (Eq 2). De plus, à la réception des messages envoyés par le véhicule CH, le capteur passerelle peut commencer le traitement du message MU envoyé par le véhicule CH. Et la préparation du message MSG envoyé à ce dernier même avant qu'il ne le détecte.

À l'inverse dans le protocole à deux phases, le capteur passerelle peut traiter les messages uniquement quand il détecte le véhicule CH.

Lorsque le véhicule est dans la portée de transmission du capteur passerelle (figure 2)

Capteur passerelle: à la détection d'un véhicule (Tdétecte), si le capteur a déjà préparé (traitement, agrégation de données, élimination de redondances et mise en format de données) le message MSG, il l'envoie directement au véhicule sinon il l'envoie dès qu'il terminera sa préparation.

Le message MSG sera enregistré dans la base de données du capteur passerelle pendant une période Tmsg pour l'envoyer à un autre véhicule (relais). Ce message est détruit si la période Tmsg est terminée ou quand le capteur passerelle reçoit l'accusé de réception du véhicule CH. Si un message critique MC arrive au capteur passerelle par un capteur détecteur, au temps où le véhicule est dans sa portée de transmission, il l'envoie directement en priorité au véhicule.

Diffusion dans VANET: après que le véhicule envoie l'accusé de réception (ACK) au capteur passerelle, il diffuse le message MSG et les messages critiques dans son groupe (toutes les entités (les véhicules, capteurs) dans le réseau peuvent recevoir ces derniers). Ensuite, il met à jour sa base de données.

Remarque : à la prochaine période de diffusion du message beacon, le véhicule CH ne diffuse que ce message, dans le cas où il se trouve dans la même zone et les mêmes étapes seront appliquées si ce message est reçu par un autre capteur passerelle.

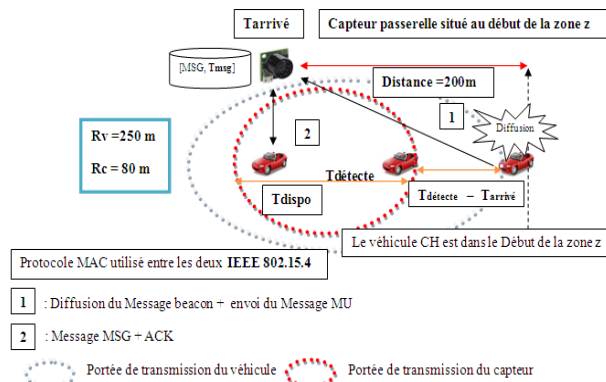


Figure 2: Echange entre le capteur passerelle et le véhicule (CH)

1.1.2 Estimation du temps d'échange des informations

Le temps disponible pour faire l'échange d'informations entre le capteur et le véhicule, est donné par l'équation (1).

Pour augmenter le temps disponible il faut diminuer le temps de détection et le temps de traitement.

Le temps de détection: le temps de détection est un temps aléatoire dépendant beaucoup du matériel utilisé, c'est-à-dire la technologie matérielle de détection utilisée dans le capteur et les conditions du réseau. Donc nous n'avons pas travaillé sur la diminution de ce dernier.

Le temps de traitement: comme nous l'avons montré dans la section III.1, nous pouvons par: l'utilisation des messages beacon, l'algorithme de traitement et le modèle de données diminuaient ce temps.

Exemple d'échange entre le capteur et le véhicule CH

Nous supposons que nous avons:

-Une région géographique contenant 256 zones où chaque zone contient un nombre variable de segments de routes.

-Un capteur passerelle qui a des informations sur les zones 0, 1, 2, 3, 4 et 5, chaque zone contient respectivement 64, 1024, 256, 128, 128, 32 segments de routes.

-Un véhicule CH entre dans la zone 2 et les zones 3, 4 et 5 qui sont dans la direction de sa destination (il est passé par les zones 0 et 1 pour arriver à la zone 2).

-Le protocole MAC considéré, par exemple, le IEEE 802.15.4 (avec un débit égal à 500 Kb/s).

a. Quantité d'informations envoyées par le véhicule CH

-Message beacon (200bits) + identificateur (20bits) + coordonnées de la destination (160 bits) + bit CH(1) + bit indicateur (1).

-Message MU informations (zone 0 et 1).-Accusé de réception (160 bits).

La quantité d'informations d'une zone donnée est calculée par l'équation 3 (la sectionII.2.2).

Donc: $Q_{tv} = Q_{z0v} + Q_{z1v} + 382 \text{ bits}$. A.N: $Q_{tv} = 43662 \text{ bits}$.

b. Quantité d'informations envoyées par le capteur passerelle

Les informations des zones 2, 3, 4 et 5

La quantité d'informations d'une zone donnée est calculée par l'équation 3 (la sectionII.2.2).

$Q_{tc} = Q_{z2c} + Q_{z3c} + Q_{z4c} + Q_{z5c}$. A.N: $Q_{tc} = 20344 \text{ bits}$.

c. Temps requis pour faire l'échange de toutes les informations

$Trequis = Q_{tv} + Q_{tc} + \text{accusé de réception} / \text{débit}$ secondes

$Tarrivé = Q_{tv} / \text{débit}$ (Tarrivé = 0,1 secondes).

La portée de transmission d'un capteur est $R_c = 80m$ et nous calculons le temps disponible d'après les trois scénarios suivants: un véhicule se déplace à une vitesse égale à 130, 100 et 60 km/h. Le Temps (Tepc) pour que le véhicule entre dans la portée de transmission du capteur, est donné par la formule suivante:

$Tepc = R_v - R_c / V$ secondes. R_v : La portée de transmission du véhicule ($R_v = 250$ mètres).

Le tableau 1 donne le temps disponible sans prendre en considération le temps de détection et de traitement par rapport à la vitesse du véhicule.

Scénario	Vitesse (km/h)	Temps disponible (secondes)	Trequis (Secondes)
L'autoroute	130	4,43	0,13
La route rurale	100	5,8	
La ville	60	9,6	

Tableau 1: Temps disponible par rapport à la vitesse du véhicule

Le tableau 2 donne le temps maximum pour le traitement des messages et la détection de véhicule par rapport à sa vitesse. Dans la solution proposée dans [2] et le protocole MEP.

Scénario	Vitesse (km/h)	Solution [2]	Protocole MEP
		Ttraitement et Tdétecte (secondes)	Ttraitement et Tdétecte (s)
L'autoroute	130	4,30	8,38
La route rurale	100	5,67	10,97
La ville	60	9,47	18,37

Tableau 2: Comparaison des temps de détection et de traitement entre la solution [2] et le protocole MEP

Dans les trois scénarios, nous avons un temps suffisant pour faire l'échange complet d'informations entre le capteur passerelle et le véhicule CH. Cet exemple nous montre l'avantage de notre proposition sur le protocole proposé dans [2] (Tableau 2). De plus dans notre solution, le capteur passerelle n'envoie au véhicule que les informations non disponibles dans sa base de données (algorithme de traitement) par contre dans la solution [2], le capteur passerelle envoie toutes les informations des segments au véhicule (le véhicule peut recevoir des informations existantes déjà dans sa base de données).

1.2 Cas B: échange entre le capteur et le véhicule (Non CH)

Véhicule (Non CH): avant que le véhicule ne diffuse un message beacon, il lui ajoute les coordonnées de sa destination, son identificateur, le bit CH (dans ce cas bit CH = 0) et le bit indicateur (bit indicateur = 0). Ensuite il diffuse ce message dans son segment.

Capteur passerelle ayant reçu ce message beacon: le capteur passerelle reçoit ce message, quand il est dans la portée de transmission du véhicule (Non CH). Il vérifie d'abord en premier, le bit CH (bit CH = 0), dans ce cas, si le message MSG est encore enregistré (Tmsg ≠ 0), il le met à jour, s'il y a de nouvelles informations (apportées par les capteurs).

Lorsque le véhicule entre dans la portée de transmission du capteur passerelle (Figure 3)

Capteur passerelle: quand il détecte le véhicule (Tdétecte). Si un message critique est arrivé au capteur passerelle par un capteur détecteur, dans le temps, où le véhicule est dans sa portée de transmission, alors il l'envoie directement en priorité à ce véhicule. Sinon il lui envoie directement le message MSG s'il est toujours enregistré.

Diffusion dans VANET: Après que le véhicule envoie l'accusé de réception (ACK) au capteur passerelle, il diffuse les messages reçus par le capteur passerelle dans son groupe, ensuite il met à jour sa base de données.

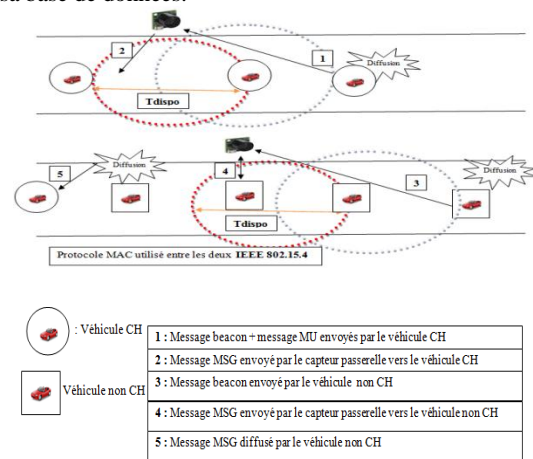


Figure 3: Echange entre le capteur passerelle et le véhicule (Non CH)

Avantages d'utilisation de véhicule non CH (comme un relais)

Le temps de traitement peut être égal à 0 ou la différence entre le temps de détection du véhicule non CH et le temps de traitement des messages MU lors de la réception du message beacon et du message MU envoyés par le véhicule CH au capteur passerelle.

Même dans le cas où le chef de groupe a reçu les messages critiques, quand il l'est diffuse dans son groupe, il y a une probabilité que quelques véhicules ne reçoivent pas ces messages (problèmes de collision et d'interférence). Mais, comme le véhicule non-CH diffuse aussi ces messages (duplication limitée) dans le groupe, ces véhicules peuvent recevoir ces derniers.

Le chef de groupe peut recevoir de nouvelles informations qui sont collectées par le capteur passerelle après avoir quitté sa portée de transmission grâce au véhicule relais.

1.3 Cas C: échange entre le capteur passerelle et le véhicule (protocole à deux phases)

Si le message beacon diffusé par le véhicule (CH ou relais) n'est pas arrivé au capteur (problème d'interférence ou de collision) ou il est arrivé après que le véhicule soit détecté par le capteur, le protocole utilise un algorithme de communication à deux phases (protocole à deux phases). De cette façon, même dans le cas où le message beacon diffusé par le véhicule, n'est pas reçu par le capteur passerelle ou il arrive en retard au capteur passerelle, la communication entre les deux réseaux peut avoir lieu. Ce qui démontre l'avantage de notre protocole par rapport aux solutions [5,6] dont leurs fonctionnements dépendent des contenus des messages beacon.

2. Les messages critiques

Pour assurer une livraison rapide (dans un temps optimal) des messages critiques envoyés par les capteurs passerelles aux véhicules, nous avons défini les trois concepts suivants:

1. Modèle de message
2. Capteurs passerelles temporaires
3. Algorithme de livraison des messages critiques

Dans les sous sections suivantes, nous allons expliquer chacun de ces concepts.

2.1. Modèle de message

Nous avons proposé un modèle de message (Figure 4) permettant la distinction entre un message critique et un message utile. Et pour éviter le stockage et la duplication inutiles des messages. Ce modèle est représenté par un entête ajouté par les capteurs passerelles au contenu du message. Cet entête contient les champs suivants: l'identificateur de capteur détecteur de l'événement, le type de message, le numéro de séquence (ce champ est incrémenté à 1 lors son passage du capteur) et la date de capture de cet événement sur la route.

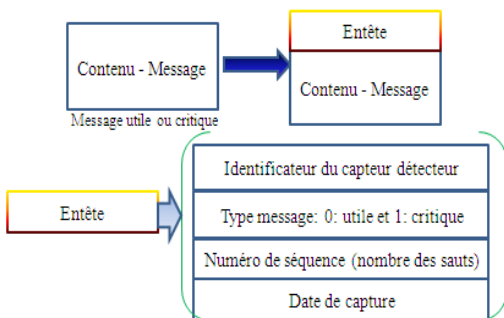


Figure 4: Modèle de message

2.2. Capteurs passerelles temporaires

Comme le nombre des capteurs passerelles est limité et les événements critiques peuvent se produire sur n'importe quel segment de route, nous avons pensé à l'utilisation des autres capteurs pour livrer les messages critiques aux véhicules. Notre proposition permet de changer temporairement le rôle d'un capteur ordinaire à un capteur passerelle uniquement dans le cas suivant: un événement critique est détecté par ce capteur au moment, où un véhicule est présent dans sa portée de transmission.

Remarque: nous avons violé la notion de groupe dans les réseaux de capteurs sans fil uniquement dans ce cas. Notre but est d'assurer une livraison rapide des messages critiques aux véhicules (le délai est la métrique la plus importante dans ce type de protocoles).

La figure 5 représente un exemple sur l'intérêt d'utilisation des capteurs passerelles temporaires pour transmettre les messages critiques aux véhicules. Dans cette figure, on remarque que la voiture rouge est informée sur la présence d'une voiture venant de l'autre sens par le capteur détecteur. De la même façon, la voiture grise est informée par un capteur de routage.

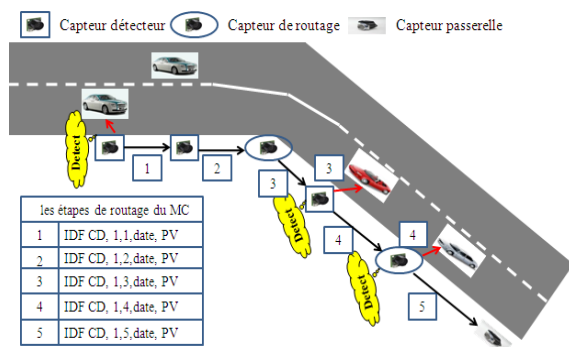


Figure 5: Exemple sur les capteurs passerelles temporaires

2.3. Algorithme de livraison des messages critiques

Nous avons proposé un algorithme de livraison des messages critiques appliqué par les capteurs passerelles pour réalisation des deux points suivants:

- L'élimination des messages critiques redondants
- Le trie de l'ordre d'envoi des messages critiques selon leur degré d'importance.

La conception de cet algorithme se base sur la lecture de l'entête ajouté au contenu des messages critiques. À travers les champs de cet entête et le codage simple de l'information critique (modèle de donnée [la section II.2.2]), le capteur passerelle peut définir les messages redondants (l'identificateur de capteur détecteur, la date de capture de l'événement et le contenu de message critique). Ainsi qu'il peut les ordonner selon leur degré d'importance (le numéro de séquence, la date de capture et le contenu de message).

La figure 6 représente un exemple sur l'utilisation de cet algorithme. Dans cet exemple, à un instant donné le capteur passerelle reçoit cinq messages critiques différents. Il applique l'algorithme de livraison pour éliminer les messages redondants et les ordonner selon leur degré d'importance. Lorsqu'il détecte le véhicule CH, il les envoie en priorité à ce dernier selon l'ordre obtenu par l'application de cet algorithme. Après l'échange des messages utiles entre les deux aura lieu. À cet instant un accident se produit à cause d'une collision entre deux véhicules. Le capteur qui a détecté cet événement, transmet un message critique de celui-ci au capteur passerelle par les capteurs de routage. À la réception de ce message par le capteur passerelle, il interrompt l'échange d'informations utiles avec le véhicule afin d'envoyer ce message à ce dernier.

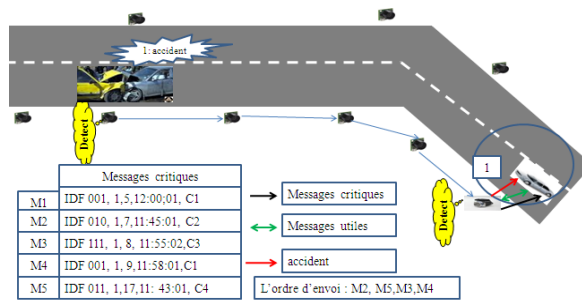


Figure 6: Exemple sur l'utilisation d'algorithme de livraison

IV. RESULTATS DE SIMULATION

Nous avons utilisé le simulateur OMNeT ++ 4.3 [6] et le projet MiXiM 2.3 [7] (pour le modelé de mobilité et le protocole MAC (IEEE 802.15.4)) pour évaluer les performances de notre Protocole (Tableau 3). L'objectif de ce protocole est d'assurer un échange total d'informations routières (messages utiles) et une transmission rapide des messages critiques aux véhicules dans le temps disponible où le véhicule est dans la portée de transmission du capteur, quelque soit sa vitesse. Pour cette raison nous avons travaillé sur:

- Les délais d'échanges des paquets entre un capteur et un véhicule.
- Le nombre maximal des paquets échangés entre un capteur et un véhicule dans le temps disponible.
- Les délais de livraison des messages critiques.

1. Environnement et paramètres de la simulation

1.1. Environnement de la simulation

1. Un réseau hybride déployé dans une région géographique contenant 1024 zones, chacune de ces zones contient un nombre variable de segments. Nous avons pris une partie du réseau hybride composée de:
2. Un chef de groupe véhiculaire a des informations sur les deux zones 0 et 1 qui ont respectivement 64 et 128 segments de routes.
3. Deux capteurs passerelles, le premier a des informations sur les zones 0,1, 2, 3 et 4, le nombre des segments des zones 2,3 et 4 sont respectivement 128,256 et 64 segments de routes. Et le deuxième capteur a des informations sur les zones 3, 4, 5 et 6, le nombre des segments dans les deux zones 5 et 6 sont respectivement 32,128 segments de routes.
4. La taille d'un accusé de réception est 160 bits.
5. La route de simulation est une route unidirectionnelle, de 4 Km de longueur. Nous supposons qu'il y a 4 véhicules se déplaçant à une vitesse variable de moyenne égale à μ Km/h sur la route. (Les 4 véhicules composent un groupe avec le premier dans la direction du capteur représente le chef de groupe véhiculaire), la distance initiale entre ces derniers est 200 mètres.
6. Les zones 3, 4, 5 et 6 sont dans la direction des véhicules vers leur destination.

Nous avons les cas d'échanges suivants:

- 1^{er} cas: échange entre le 1^{er} capteur et le véhicule CH.
- 2eme cas: échange entre le 2eme capteur et le véhicule CH.
- 3eme cas: échange entre le 1er capteur et le véhicule non CH.
- 4eme cas: échange entre le 2eme capteur et le véhicule non CH.

1.2. Paramètres de la simulation

- Pour les cas 1, 2, 3 et 4, nous avons les totaux des tailles des messages qui sont respectivement: 3050, 3812, 2166 et 2307 octets.
- Le capteur passerelle ne détruit pas les paquets envoyés au véhicule CH, même s'il a reçu l'accusé de réception de ce dernier (nous avons ajouté cette hypothèse pour voir l'importance de l'utilisation d'un véhicule non CH comme un relais).

Paramètre	Valeur
Portée de transmission d'un capteur R_s	80 mètres
Portée de transmission d'un véhicule R_v	250 mètres
Nombre des nœuds de capteurs	2
Nombre des véhicules	4
Vitesse moyenne des véhicules μ	Entre 40 à 140 km heure
Protocole MAC	IEEE 802.15.4
Débit de transmission	500 kilo bits seconde
Les tailles des paquets	Entre 20 et 100 octets

Tableau3: Paramètres de simulation

2. Résultats et interprétation

2.1 Les délais des échanges des paquets

L'objectif principal de notre protocole est le délai d'échange des paquets entre le capteur et le véhicule par rapport à des vitesses variables du véhicule. Ce délai peut monter si le protocole assure l'échange de données dans le temps disponible (où le véhicule est dans la portée de transmission du capteur). Nous avons testé le protocole dans les deux situations suivantes:

1^{ere} situation: nous supposons que les messages beacon diffusés par le véhicule arrivent au capteur avant qu'il ne détecte le véhicule (Protocole MEP avec l'utilisation de message beacon). (Figure 7).

2eme situation: nous supposons que le message beacon n'arrive pas au capteur ou il arrive après que le capteur a détecté le véhicule ayant envoyé ce dernier. (Protocole MEP à deux phases [1,2,3]). (Figure 8).

Interprétation générale des résultats

Notre protocole assure l'échange d'informations dans le temps disponible dans les deux cas (avec/sans l'utilisation des messages beacon). Si nous prenons le cas où la vitesse d'un véhicule est égale à 140 km/h (vitesse maximale). Nous avons le temps disponible suivant: $T_{dispo} = 4,11$ secondes qui est supérieur aux temps obtenus dans la simulation (Figures 7 et 8) à la même vitesse. Nous avons utilisé les deux approches (la première se base sur l'utilisation de message beacon et la deuxième se base sur la communication à deux phases) et nous préférons la première approche pour les raisons suivantes:

Les délais des échanges d'informations entre le capteur et le véhicule (Figure 8 [2]) sont plus grands que les délais obtenus dans le cas de l'utilisation des messages beacon (Figure 7). Ce qui justifie notre choix concernant l'utilisation de message beacon dans notre protocole en priorité. S'il y a des interférences, nous passons à l'utilisation de connexion à deux phases pour éviter le temps supplémentaire dû au retard de l'arrivée du message beacon.

Dans le cas d'interférence ou de forte collision (le message beacon n'est pas reçu par le capteur) nous pouvons utiliser la deuxième approche. Ce qui démontre la fiabilité de notre protocole par rapport aux autres protocoles se basant sur les messages beacon [5,6], c'est-à-dire si le capteur n'a pas reçu ce message, ces protocoles peuvent avoir des problèmes de fonctionnement.

2.2 Le nombre maximal des paquets échangés

Dans ce cas, nous avons fait la simulation du protocole pour voir le nombre maximal des paquets échangés entre les capteurs et les véhicules dans le temps disponible d'échange. (Figures 9 et 10). Nous avons utilisé des paquets de taille fixe égale à 100 octets (en général, chaque paquet contient des informations sur des segments des routes et qui sont destinées au chef de groupe (véhicule N°1)).

Interprétation générale des résultats

L'utilisation de véhicule (Non CH) comme un relais, permet dans plusieurs cas de récupérer des paquets qui n'ont pas été reçus par le véhicule CH (Figures 9 et 10).

La figure 11, montre le taux moyen de duplication des paquets envoyés au véhicule CH dépendant du choix de la période T_{msg} . Le choix de cette période influe sur l'efficacité du protocole, il dépend de plusieurs facteurs qui sont les suivants:

1. L'environnement (autoroute, ville et route rurale) 2. La vitesse moyenne de déplacement des véhicules 3. La densité des véhicules.

Dans un même environnement cette période peut être changée par rapport aux informations apportées (la densité des véhicules, le taux de collision...) par les capteurs.

La figure 12 montre le taux des paquets perdus, nous avons obtenu un taux faible de perte des paquets. Ce qui démontre la fiabilité des algorithmes utilisés dans le protocole.

2.3 Les délais de livraison des messages critiques

Nous supposons que:

- Le 1^{er} capteur n'est pas un capteur passerelle et le 2eme capteur est un capteur passerelle.
- Il y a cinq messages critiques qui sont transmis au deuxième capteur.
- Les 4 premiers messages sont transmis par des capteurs détecteurs qui ont détectés les événements critiques dans des temps différents. Le 5eme message est transmis par le premier capteur au deuxième capteur.

Message1: (000000010 001,1, 0011,00 :01 :23) + Contenu message

Message2: (000000010 011,1, 0100, 00 :01 :27) + Contenu message

Message3: (000000010 111,1, 1100, 00 :01 :30) + Contenu message

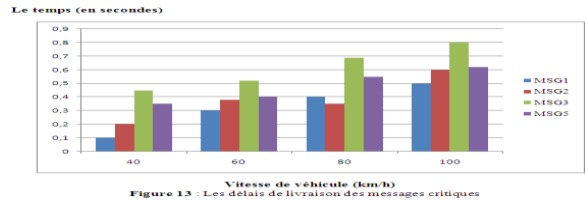
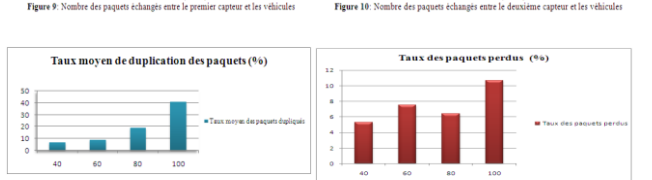
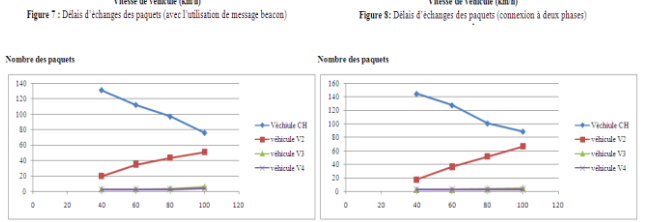
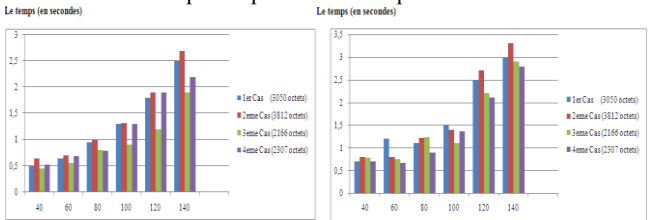
Message4:(000000010 001,1, 1010, 00 :01 :29) + Contenu message (le même contenu du message 1)

Message5:(000000010 010,1, 0001, 00 :01 :28) + Contenu message

Utilisation d'un capteur passerelle temporaire (1^{er} capteur)

Pour le temps d'arrivée du message 5 (du 1^{er} capteur au véhicule CH), nous avons obtenu les résultats suivants: **0.15, 0.23, 0.3 et 0.52** secondes respectivement aux vitesses suivantes: 40, 60, 80 et 100 km/h.

La figure 13 montre les délais de livraison des cinq messages, par rapport à des vitesses variables du véhicule. Nous remarquons que le message 4 est éliminé par l'application de l'algorithme de livraison et les temps d'arrivées des messages sont ordonnés selon leur degré d'importance. De plus nous avons obtenu des délais de livraison courts (tous les délais < 1 seconde) par rapport au temps disponible. L'utilisation de 1^{er} capteur pour transmettre le message 5, permet de réduire les temps d'arrivées de ce message par rapport aux temps d'arrivées de même message par l'utilisation de 2eme capteur (capteur passerelle).Ce qui démontre l'intérêt d'utilisation des capteurs passerelles temporaires.



V.CONCLUSION

Dans cet article, nous avons présenté un protocole de communication entre un WSN et un VANET dans les HSVN. Ce protocole utilise des solutions permettant l'amélioration de son rendement. Nous pouvons résumer ces solutions dans les points suivants: un fonctionnement hybride du protocole (en utilisant les messages beacon, connexion à deux phases) par rapport aux conditions du réseau, une réduction de traitement des messages utiles (le modèle simple de données, l'algorithme de traitement utilise par le capteur), une communication unique entre le capteur passerelle et le chef de groupe véhiculaire pour économiser l'énergie du capteur, une utilisation de véhicule relais dans le cas où l'échange entre le capteur passerelle et le véhicule CH ne se termine pas avec succès, un modèle de message, un algorithme de livraison des messages critiques selon leur degré d'importance et finalement des capteurs passerelles temporaires. Nos perspectives sont l'amélioration de notre protocole afin qu'il s'adapte avec d'autres environnements et par l'ajout d'autres métriques dans la simulation comme la consommation d'énergie du capteur pour plus fiabilité.

REFERENCES

- [1] Kong, F.; Tan, J. A Collaboration-Based Hybrid Vehicular Sensor Network Architecture. In *Proceedings of the International Conference on Information and Automation*, Zhangjiajie, China, 20–23 June 2008; pp. 584-589.
- [2] Tripp, C.; Ornelas, K.; Aguilar, M. Performance Evaluation of a Hybrid Sensor and Vehicular Network to Improve Road Safety. In *Proceedings of the ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks*, Bodrum, Turkey, 17–18 October 2010; pp. 1-8.
- [3] Andreas Festag, Alban Hessler, Roberto Baldessari, Long Le, Wenhui Zhang, Dirk Westhoff. NEC Laboratories Europe "Network Research Division", Kurfürsten-Anlage 36, D-69115 Heidelberg "VEHICLE-TO-VEHICLE AND ROAD-SIDE SENSOR COMMUNICATION FOR ENHANCED ROAD SAFETY" 2010.
- [4] JianFeng, L and Jianglong, H :A RESTful information service method in Hybrid Sensor and Vehicular Networks publisher in *Automatic Control and Artificial Intelligence (ACAI 2012), International Conference* Xiamen- China 3-5 March 2012, pp. 283 – 286.
- [5] Weingärtner, E.; Kargl, F. A Prototype Study on Hybrid Sensor-Vehicular Networks. In *Proceedings of the 6th GIITG KuVS Fachgespräch "Wireless Sensor Networks"*, Aachen, Germany, 16–17 July 2007. K. Elissa.
- [6] S.djahal and Y. Hamri_doudane ENSIIE :A Framework for Efficient Communication in Hybrid Sensor and Vehicular Networks publisher in "the 9th annual IEEE consumer Communication and Networking Conference ,CCNC 1214-17 Jan. 2012,pp. 209 – 214.
- [7] OMNeT++ Community - www.omnetpp.org. 2001-2009.
- [8] <http://apps.sourceforge.net/trac/mixim.html> Generated on Mon Nov 21 2011 06:38:23 for MiXiM by doxygen 1.7.5.

Un protocole de négociation pour l'ordonnancement distribué Multi-Agents d'un Atelier Job-Shop

M.L. Berrandjia, S. Ourari

Division Robotique et Productique
Centre de Développement des Technologies Avancées
Alger, Algérie
mberrandjia@cda.dz

R. Chalal

Laboratoire de Méthodes de Conception des Systèmes
Ecole Nationale Supérieure d'Informatique
Alger, Algérie
r_chalal@esi.dz

Résumé—Dans ce papier, nous nous intéressons à la résolution distribuée du problème d'ordonnancement d'atelier Job-Shop en utilisant les systèmes multi-agents. Nous considérons que chaque machine est assimilée à un agent autonome gérant son propre ordonnancement local (Ordonnancement à une machine) et coopérant avec les autres agents pour trouver une solution globale satisfaisante tout en gardant une certaine flexibilité séquentielle afin de faire face aux incertitudes. L'approche de résolution proposée est basée sur un mécanisme de coopération inter-ressources où les négociations sont initiées entre les agents afin d'assurer une cohérence entre tout couple de tâches liées par la gamme opératoire. Un protocole de négociation est proposé et sa mise en œuvre est réalisée sur la plateforme de développement des systèmes multi-agents JADE. Cette dernière permet de développer une solution portable fonctionnant sur plusieurs plateformes et supportant le standard FIPA-ACL pour la communication entre agents.

Mots clés—*Job-Shop, ordonnancement distribué, système multi-agents, coopération, flexibilité, JADE.*

I. INTRODUCTION

Actuellement, le processus d'ordonnancement de la production occupe une place importante au niveau de l'entreprise qui évolue dans un environnement caractérisé par une hyper-concurrence visant à satisfaire à la fois les exigences et les contraintes imposées par les clients en termes de qualité, de coût et de délais de mise à disposition. Il est donc indispensable d'avoir un système assurant un ordonnancement robuste tenant compte de l'environnement perturbé de la production en exploitant au maximum la flexibilité existante lors de l'exécution du plan de production.

Il existe dans la littérature plusieurs approches traitant le problème d'ordonnancement de la production. La plupart de ces approches sont centralisées et s'articulent essentiellement autour de méthodes de résolution exactes ou bien approchées [1]. La solution est alors donnée sous forme d'un ordonnancement prévisionnel unique satisfaisant les contraintes imposées par le client ainsi que par l'environnement de l'atelier. Notons toutefois, que le principal inconvénient de ces approches réside dans le fait que ces dernières considèrent que les problèmes d'ordonnancement sont déterministes, i.e. que les données du problème traité sont connues à l'avance. Seulement dans la réalité, l'environnement de production est de nature perturbé; en effet de nouvelles informations peuvent surgir pendant l'exécution du plan initial le rendant par la même occasion obsolète. La solution dans ce cas consiste donc à élaborer un nouvel ordonnancement global couteux en temps de calcul conduisant le plus souvent à des modifications

radicales du plan initial ou bien dans le pire des cas à l'acceptation du retard engendré par la perturbation.

Pour faire face au problème d'incertitudes, des approches dites robustes ont été développées. Ces approches sont classées en trois catégories : prédictives, proactives et réactives. Dans les approches prédictives, l'ordonnancement est calculé en se basant sur des données estimées sans tenir compte des perturbations. Les approches proactives quand à elles calculent un ordonnancement en utilisant des connaissances a priori sur les incertitudes probables. Enfin, les approches réactives effectuent un calcul de l'ordonnancement en temps réel en traitant les incertitudes lors de leur apparition. Il à noter que les approches précédentes peuvent être combinées afin d'exploiter les avantages que chacune de ces dernières offre. On retrouve dans [2] et [3] un état de l'art et une classification des approches robustes.

D'autres approches exploitent la nature distribuée du problème d'ordonnancement, les décisions sont dans ce cas distribuées entre les différents acteurs, ayant chacun une autonomie, et coopérant entre eux de manière à converger vers un ordonnancement global avec une performance acceptable. Cette façon de faire permet de mieux absorber les incertitudes en ne modifiant que les ordonnancements locaux des ressources concernées par la perturbation. Notons que ce type d'approche distribuée peut être modélisé par les systèmes multi-agents (SMA) [4] en considérant que chaque agent représentant une ressource (machine) possède une autonomie décisionnelle et est capable de communiquer avec les autres agents moyennant un protocole de coopération pour la prise de décisions d'ordonnancement. L'intérêt de ces SMA s'est accru ces dernières années avec le développement rapide des réseaux et de l'informatique distribuée. Plusieurs chercheurs se sont intéressés au paradigme des systèmes multi-agents et à leur application dans le domaine de l'ordonnancement de la production. Dans [5], les auteurs décrivent un mécanisme d'appel d'offre permettant à une ressource de sous-traiter ses tâches avec d'autres ressources. Baker dans [6] a proposé une architecture basée sur le protocole Contract-Net. Dans [7], le système développé assigne à chaque ressource un agent permettant de forcer le respect des contraintes relatives à la capacité de cette dernière, et à chaque travail un agent s'assurant du respect des contraintes de précédence et des dates de disponibilités. Dans les articles [8] et [9], l'ordonnancement est réalisé avec deux agents qui détiennent un ensemble de tâches et sont en compétition pour l'utilisation des ressources. Dans [10], le problème d'atelier job-shop a été abordé dans un contexte coopératif et distribué pour la prise en compte des incertitudes. Dans [11], les auteurs présentent un SMA pour l'ordonnancement des

ateliers job-shop où les agents définissent leurs plans de réalisation en utilisant des règles de priorité tout en minimisant les temps d'inactivité de la machine.

Dans ce papier, nous nous intéressons à la résolution distribuée d'un problème d'ordonnancement multi-ressources et plus particulièrement à sa mise en œuvre en utilisant la plateforme JADE (Java Agent DEvelopment Framework) [12]. Cette dernière est organisée selon les spécifications de la norme FIPA (Foundation for Intelligent Physical Agents) appelée FIPA97 [13] qui décrit le modèle de référence d'une plateforme multi-agents ainsi que le langage de communication entre les agents (FIPA-ACL).

Ce papier est organisé comme suit. Dans la section suivante, nous présentons la problématique d'ordonnancement étudiée. La troisième section est consacrée à la présentation de l'approche multi-agents proposée et une description du protocole de négociation y affèrent. Par la suite, nous décrivons la mise en œuvre du protocole proposé pour l'ordonnancement multi-agents en utilisant la plateforme JADE. Enfin, nous terminons par une conclusion sur le travail réalisé et les perspectives futures en vue.

II. PROBLEMATIQUE D'ORDONNANCEMENT JOB-SHOP

Dans ce papier, nous étudions le problème d'ordonnancement de type Job shop noté J_n/C_{max} et qui consiste à ordonnancer un ensemble de travaux $T = \{1, 2, \dots, N\}$ sur un ensemble de machines $M = \{1, 2, \dots, m\}$. Chaque travail possède sa propre gamme opératoire et est composé d'un ensemble de tâches s'exécutant sur une des machines de l'ensemble M dans un ordre bien défini. Nous supposons que les tâches sont non-préemptives, et que les machines sont disjonctives. Notons aussi que chaque tâche i ne peut être exécutée que par une et une seule machine et que son exécution nécessite une durée opératoire p_i connue à l'avance.

La solution du problème consiste à trouver pour chaque machine un ordonnancement de manière à satisfaire les contraintes du problème et à converger au mieux vers un objectif qui est la minimisation de la durée totale de la réalisation des N travaux appelée Makespan et notée C_{max} . Ce genre de problème est connu comme étant classé dans la catégorie des problèmes NP-Difficiles [14].

Se situant dans un contexte perturbé, nous exploitons, pour résoudre le problème, la nature distribuée de la fonction ordonnancement, en considérant chaque machine comme étant un centre de décision autonome disposant de toutes les informations nécessaires (connaissances locales) relatives aux tâches (date de disponibilité, date d'échéance et durée opératoire) devant être exécutées sur ce dernier. Ces connaissances locales sont utilisées par le centre de décision pour réaliser son propre ordonnancement local en définissant les dates de début des ses opérations. Le problème qui se pose alors est la cohérence des décisions prises par les différents centres de décisions sachant qu'ils sont interdépendants en raison des liens qui relient les différentes tâches d'un même produit (une date de fin d'une tâche ne pouvant excéder la date de début de la tâche qui la succède). En effet, en supposant que chaque centre prend ses décisions sans prendre en compte celles prises par les autres centres, il est clair que cela engendrerait des incohérences qu'il faudrait éliminer par la suite. On peut déjà en conclure que ces centres de décision sont dans l'obligation d'interagir les uns

avec les autres pour assurer la cohérence entre leurs décisions et cela en utilisant un mécanisme de coopération bien précis. Un centre décision i possède des relations avec des centres en amont qui lui fournissent des produits à transformer, et qu'il il transfère à son tour aux centres de décision en aval chargés de la suite des opérations.

Parmi les paradigmes les plus utilisés pour la résolution des problèmes distribués, les systèmes multi-agents (SMA) en constituent le moyen le plus adéquat pour la modélisation distribuée du problème d'ordonnancement en offrant une multitude d'outils et de standards tels que ACL ou bien KQML.

III. APPROCHE MULTI-AGENTS POUR L'ORDONNANCEMENT

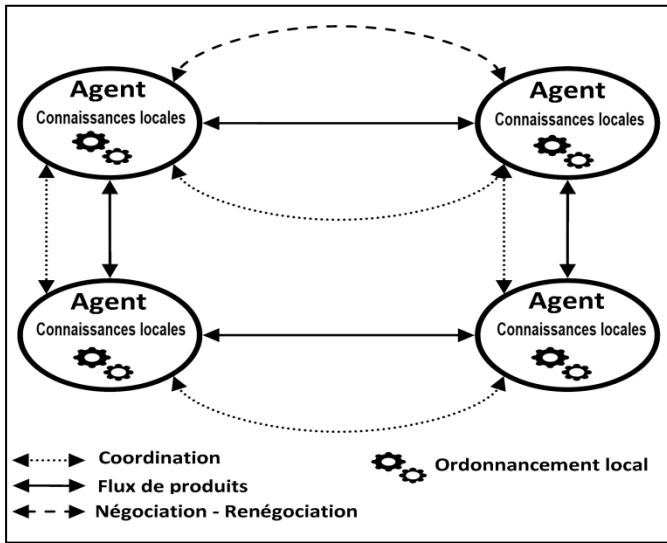
Comme nous l'avons introduit précédemment, nous considérons dans l'approche proposée pour la résolution distribuée du problème, que chaque machine est assimilée à un agent, et que ce dernier procède à l'établissement de son ordonnancement local en exploitant la technique d'ordonnancement à une machine basée sur le théorème de dominance décrite dans [15] et qui permet de caractériser un ensemble de solutions faisables au lieu d'une solution unique et fournit ainsi de la flexibilité séquentielle. Ainsi, chaque agent, définit pour chaque tâche i qu'il réalise, des dates de début et de fin au mieux et au pire et dispose d'une certaine flexibilité lui permettant de faire face aux incertitudes. Notons que chaque ordonnancement local déterminé par un agent donné est réalisé indépendamment des décisions prises localement par les agents en amont et en aval de ce dernier, engendrant ainsi des incohérences qu'il faudrait éliminer afin d'aboutir en finalité à un plan d'ordonnancement global réalisable. Afin d'assurer une cohérence entre les décisions prises localement, un mécanisme de coopération entre les agents est mis en œuvre à travers des primitives, permettant ainsi de négocier de nouvelles valeurs de début et de fin en jouant sur les dates de disponibilités et d'échéances des tâches. La phase de négociation est initiée par les deux agents ayant la plus grande incohérence. Pendant cette phase, les agents utilisent des primitives de négociation leurs permettant de faire des propositions ou bien des contre-propositions afin d'aboutir à la fin à une situation sans incohérences. Le ré-ordonnancement effectué au niveau d'un agent négociant nécessite une mise à jour des connaissances au niveau des autres agents à travers une coordination. Il est à noter qu'un couple d'agents peut faire l'objet de plusieurs négociations. On appelle donc une négociation déjà faite, une renégociation.

Fig. 1. Relations entre les agents ressources.

A. Décomposition du Problème Initial

Cette étape consiste à décomposer le problème d'ordonnancement job shop initial à plusieurs machines en plusieurs sous-problèmes à une machine. Il s'agit de déterminer une connaissance locale initiale en calculant pour chaque tâche i à exécuter par une ressource donnée, sa fenêtre d'exécution $[r_i, d_i]$ (r_i, d_i représentent respectivement la date de disponibilité et la date d'échéance de la tâche i). Il faut noter que les données initiales du problème concernent seulement le nombre de travaux à réaliser, leurs gammes opératoires ainsi que les ressources, et les durées opératoire des tâches à exécuter. Les dates r_i et d_i d'une tâche donnée sont calculées en utilisant le graphe disjonctif correspondant au problème [16]. Ainsi, chaque ressource (agent) disposera d'un ensemble de tâches à réaliser avec toutes les

informations (r_i , d_i et p_i) nécessaires pour procéder au calcul d'un ordonnancement local.



B. Ordonnancement local

Le but de l'ordonnancement global étant la minimisation du Makespan, le problème job shop est décomposé en sous-problèmes plus faciles à résoudre, où chaque agent résout un sous-problème, et procède à l'établissement d'un ordonnancement local à une machine avec comme critère la minimisation du retard algébrique (problème noté $1|r_i|L_{max}$). Comme nous supposons que chaque agent caractérise un ensemble d'ordonnements possibles, ceci est réalisé en exploitant une condition de dominance [15] à travers les fenêtres temporelles d'exécution des tâches $[r_i, d_i]$ calculées précédemment. Il est possible ainsi de déterminer pour chaque tâche i ce qui suit :

- La séquence favorable (resp. défavorable) avec le meilleur retard algébrique noté L_i^{min} (resp. L_i^{max});
- La date de début au mieux (resp. au pire) notée s_i^{min} (resp. s_i^{max});
- La date de fin au mieux (resp. au pire) notée f_i^{min} (resp. f_i^{max}).

Le calcul de L_i^{min} et de L_i^{max} est effectué en utilisant l'algorithme décrit dans [17] dont la complexité temporelle est en $O(n \log n)$. Les dates de début et de fin au mieux et au pire sont déduites des retards algébriques.

C. Détermination des incohérences

Dans la section précédente, nous avons montré que des intervalles temporelles spécifiant les dates de début et de fin au mieux et au pire de chaque tâche sont déterminés localement. Il faut toutefois noter que chaque agent procède à l'établissement de son ordonnancement local sans tenir compte des décisions prises par les autres agents avec qui il est lié au sens de la gamme opératoire des produits. Il en

découle ainsi des incohérences entre les dates de début et de fin des tâches, dues aux contraintes de précédence liées aux gammes opératoires. En effet, les intervalles temporelles de début de réalisation déterminés par deux agents prenant en charge respectivement deux tâches successives i et $i+1$ d'un même produit doivent satisfaire les deux conditions suivantes :

$$s_{i+1}^{min} \geq f_i^{min} \text{ et } s_{i+1}^{max} \geq f_i^{max} \quad (1)$$

Les incohérences au mieux et au pire, entre deux tâches (i , $i+1$) sont données par les formules suivantes

$$\Delta_i^{min} = s_{i+1}^{min} - f_i^{min} \text{ et } \Delta_i^{max} = s_{i+1}^{max} - f_i^{max} \quad (2)$$

Ainsi, chaque agent gère deux ensembles de tâches incohérentes, au mieux et au pire, classés respectivement par ordre croissant des Δ_i^{min} et Δ_i^{max} . Le but étant d'arriver à trouver un ordonnancement global satisfaisant les conditions données dans (1), il est clair que seules les valeurs négatives des expressions données dans (2) constituent des incohérences, et feront par conséquent l'objet de négociations entre les agents. Notons que les agents se coordonnent par échange d'informations afin que le couple d'agents détenant l'incohérence la plus élevée (la plus petite valeur négative) enclenche le processus de négociation. Nous signalons que dans le protocole de négociation que nous proposons, à un instant donné, seul un couple d'agents concernés est en phase de négociation, les autres agents entrent en état d'attente jusqu'à ce que la négociation prenne fin.

D. Protocole de négociation

Afin d'assurer une cohérence lors des prises de décisions locales, les agents interagissent selon un schéma de coopération que nous décrivons dans cette section. Comme nous l'avons déjà signalé, il s'agit plus précisément d'amorcer une série de propositions et de contre-propositions entre les agents du couple ayant la plus grande incohérence. Pour ce faire, des primitives permettant d'agir sur les dates de début et de fin des tâches au mieux et au pire sont définies au niveau des agents, permettant à ces derniers de prendre des décisions locales, tout en négociant avec les agents amonts et aval. Dans ce qui suit, nous présentons ces primitives puis nous décrivons le schéma d'interaction adopté.

1) *Primitives de gestion des incohérences*: chaque agent dispose de deux types de primitives qu'il gère de manière autonome. La première règle les incohérences au mieux, alors que la seconde traite le cas des incohérences au pire. Le principe consiste à modifier les dates de début ou de fin au mieux et au pire de manière à ce que les valeurs Δ_i^{min} et Δ_i^{max} soient positives ou bien nulles, assurant ainsi la cohérence entre les tâches. Ces primitives agissent sur l'ordre de certaines tâches déjà préétabli en augmentant les dates de début ou bien en réduisant les dates de fin (au mieux et au pire) de la tâche faisant objet de l'incohérence la plus grande. Ces primitives sont :

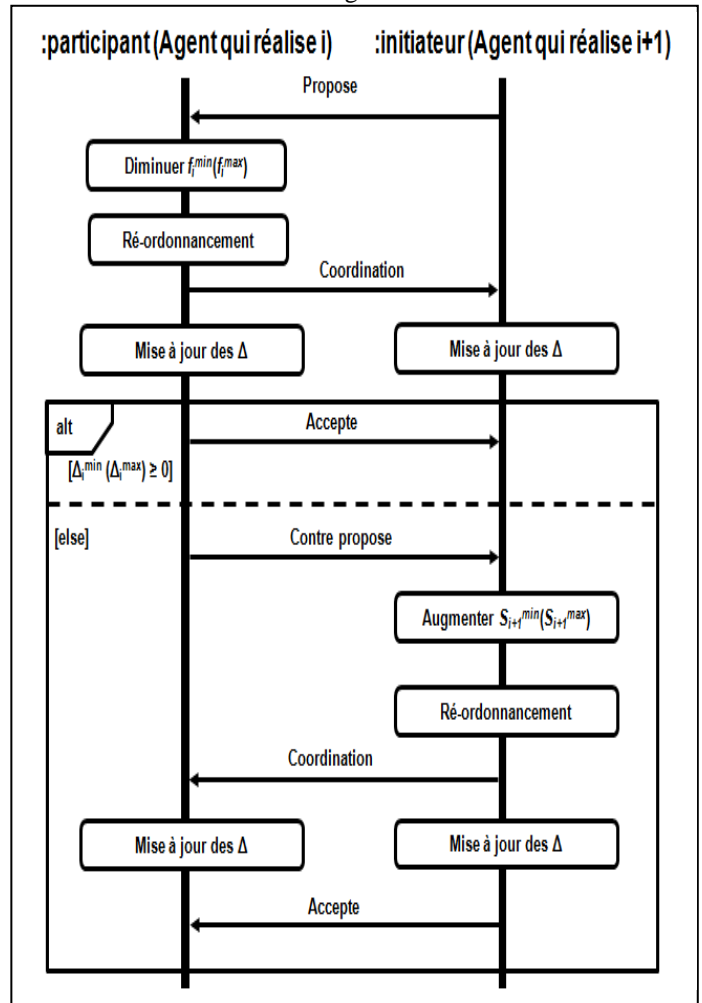
- *Diminuer* f_i^{min} (resp. f_i^{max}): cette primitive permet de diminuer la valeur de la date de fin au mieux (resp. au pire) de la tâche i ;
- *Augmenter* s_i^{min} (resp. s_i^{max}) : cette primitive permet d'augmenter la valeur de date de début au mieux (ou au pire) de la tâche i .

2) *Schéma d'interaction* : notre approche de résolution préconise, comme nous l'avons déjà signalé, la distribution des décisions sur un processus d'interaction en utilisant les SMA. En effet, une solution SMA sans interaction n'est guère différente d'une solution classique. Elle consiste à mettre en relation plusieurs agents afin de déclencher certaines actions au niveau de ces derniers. Dans notre cas, deux agents entrent en négociation pour éliminer l'incohérence liée aux tâches successives ($i, i+1$) de la même gamme, que chacun d'eux réalise. Durant cette phase de négociation, ces agents adoptent deux rôles différents ; l'agent réalisant la tâche $i+1$ est l'agent *initiateur* de la négociation alors que celui réalisant la tâche i est l'agent *participant*. Cette négociation se fait au moyen des primitives suivantes :

- *Propose* : cette primitive permet à l'agent initiateur de demander à l'agent participant de diminuer sa date de fin au mieux ou au pire ;
- *Contre-proposition* : cette primitive est utilisée par l'agent participant pour demander à l'agent initiateur d'augmenter sa date de début au mieux ou au pire, dans le cas où la diminution de la date de fin n'a pas suffi à éliminer l'incohérence ;
- *Accepte* : cette primitive est utilisée par les agents négociants afin de confirmer l'acceptation de la proposition émanant du participant ou de la contre proposition émanant de l'initiateur ;
- *Coordination* : cette primitive est utilisée par les agents négociants pour informer les autres agents des modifications éventuelles opérées à leur niveau.

La figure 2 représente un diagramme de séquence illustrant la négociation entre deux agents. La négociation commence par une proposition de l'agent initiateur (Agent réalisant la tâche $i+1$) demandant à l'agent participant de diminuer f_i^{min} (ou bien f_i^{max}). Ce dernier procède à un ré-ordonnement local et informe les autres agents des modifications effectuées afin que ces derniers les prennent en considération. Dans le cas où l'action de l'agent participant a permis de régler le problème, alors il envoie un message d'acceptation à l'agent initiateur et la négociation est terminée. Dans le cas contraire, l'agent participant envoie à l'agent initiateur une contre-proposition lui demandant d'augmenter son s_{i+1}^{min} (ou bien s_{i+1}^{max}). Ce dernier après avoir procédé à un ré-ordonnement local, informe les

autres à travers un message de coordination des



modifications faites à son niveau. A l'issue de cette étape un message d'acceptation sera envoyé à l'agent participant marquant ainsi la fin de la négociation.

Fig. 2. Schéma d'interaction lors de la négociation.

IV. MISE EN ŒUVRE

Dans cette section, nous nous intéressons à la mise en œuvre de la solution SMA proposée en mettant l'accent sur la plateforme utilisée (JADE) pour la modélisation comportementale des agents, la communication et la gestion des messages d'échange entre ces derniers. Il existe plusieurs plateformes logicielles permettant soit, la simulation des systèmes multi-agents, ou bien le développement et la mise en œuvre de ces derniers. Dans [18], les auteurs présentent une étude comparative des plateformes de développement des SMA les plus utilisées.

A. Présentation de la plateforme JADE

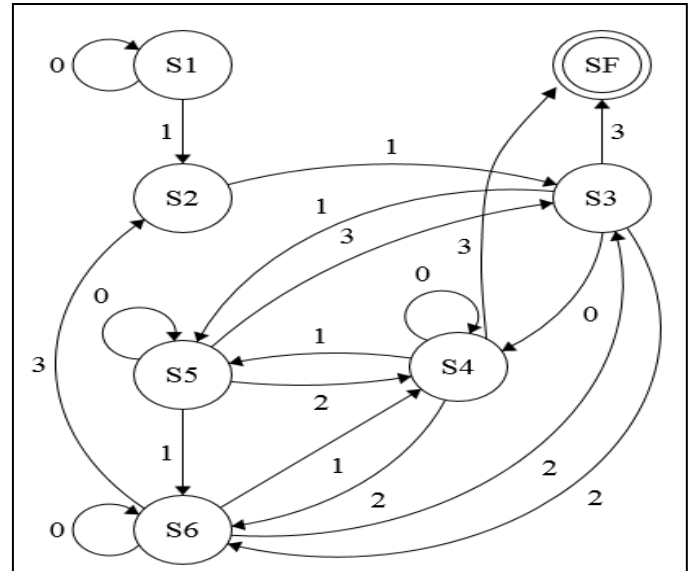
JADE (Java Agent DEvelopment framework) est une plateforme Open source gratuite, implémentée entièrement en JAVA et dédiée au développement des SMA conformément aux spécifications de la FIPA. Cette norme concerne aussi la communication inter-agents effectuée à travers le langage FIPA-ACL. JADE comporte aussi un ensemble d'outils et d'applications graphiques qui permettent le débogage, la supervision et l'administration du système développé, tels que : « *Sniffer Agent* » pour la visualisation des messages ACL circulants entre les agents du système, ou encore l'outil « *Remote Monitoring Agent* » qui permet l'administration à distance d'une ou de plusieurs plateformes JADE.

B. Modélisation multi-agent du problème sous JADE

Comme mentionné précédemment, chaque agent est autonome, possède des connaissances locales, et interagit avec les autres agents afin de mettre en cohérence les décisions prises localement. Au niveau de la plateforme JADE, chaque agent peut avoir un ou plusieurs réactions, selon la situation à laquelle il est confronté. La modélisation du comportement d'un agent est réalisée avec la classe *Behaviour* (offerte par JADE) à travers l'utilisation d'une machine à états finis (*Finite State Machine*). La figure 3 décrit le comportement général d'un agent, où chaque état correspond à une situation donnée :

- **Etat S1 (réception des données) :** durant cet état l'agent reçoit les informations sur les tâches à exécuter nécessaires pour l'établissement de son ordonnancement local. A la fin de la réception des données, l'agent passe à l'état S2 ;
- **Etat S2 (Ordonnancement local) :** cet état permet à l'agent d'établir son ordonnancement local en définissant pour chaque tâche, ses dates de début et de fin au mieux et au pire. L'agent passe ensuite à l'état S3 pour la détermination des incohérences ;
- **Etat S3 (Détermination des incohérences) :** Il s'agit dans cet état du calcul ou de la mise à jour des incohérences à chaque fois que l'ordonnancement local est établi, suite à une négociation ;
- **Etat S4 (Détermination du couple d'agents négociants) :** Cet état permet d'identifier le couple d'agents déclenchant la négociation. Ces deux derniers entrent en négociation (état S6) tandis que le reste des agents passent en attente (état S5) ;
- **Etat S5 (attente) :** cet état est relatif aux agents qui ne sont pas concernés par la négociation. Si la négociation est finie alors les agents en attente passent à l'état S4 pour déterminer le nouveau couple d'agents négociants, sinon à l'état S3 pour une coordination et une mise à jour des incohérences ;
- **Etat S6 (Négociation) :** durant cet état, les agents ayant la plus grande incohérence négocient afin d'éliminer cette dernière. Le passage à l'état S2 se fait pour effectuer un ré-ordonnancement, à l'état S3 pour mettre à jour les incohérences (coordination) ou bien à l'état S4 après la fin de la négociation.

Fig. 3. Machine à états finis d'un agent.



C. Gestion de la synchronisation inter-agents

Dans l'approche de résolution proposée, les agents sont distribués et interconnectés à travers un réseau informatique, où on ne dispose d'aucun contrôle sur l'acheminement et le routage des messages, et donc sur l'ordre d'arrivée des messages. D'un autre côté, on note que les agents s'exécutent à des vitesses différentes selon l'environnement où ces derniers sont hébergés. Ces deux derniers points posent le problème de synchronisation entre les agents. Il est donc important de procéder à une bonne gestion de la réception et de l'utilisation des messages reçus par chaque agent.

Signalons que dans la plateforme JADE, chaque agent peut envoyer des messages d'une manière asynchrone. Il dispose également d'une file d'attente lui permettant la réception asynchrone de la part des autres agents (figure 4).

Afin d'assurer un fonctionnement correct du système, une bonne gestion de la file d'attente des messages doit prendre en considération le problème de synchronisation. En effet, durant son activité, l'agent passe par plusieurs états et pendant chaque état la lecture du message approprié est importante afin régler le problème de synchronisation lié aux deux points évoqués précédemment (Vitesse d'exécution et ordre d'arrivée des messages).

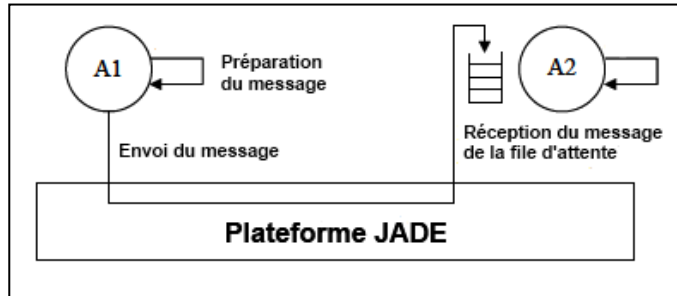
Fig. 4. Envoi et réception asynchrone des messages.

Pour pallier au problème de synchronisation et afin de permettre à chaque agent de lire le bon message au bon moment, nous avons défini pour chaque état, un type de message qui lui correspond. Il s'agit ici d'ajouter un entête permettant d'identifier le type de message afin que l'agent puisse l'extraire à partir de la file d'attente et l'utiliser, ou bien de le placer en fin de la file d'attente et passer au message suivant s'il ne correspond pas à son état actuel. A partir de cela, nous avons défini les types de messages suivants :

- **RCPT :** désigne la réception des données relatives aux tâches et la décomposition du problème initial. Ce type n'est utilisé que dans l'état S1 (réception des données) ;
- **DINC :** message relatif aux communications inter-agents visant le calcul des incohérences. Ce type

n'est utilisé que dans l'état S3 (Détermination des incohérences);

- **DCAN** : une catégorie de message associée à la phase de détermination du couple ayant la plus grande incohérence. Cette catégorie n'est utilisée que dans
- l'état S4 (Détermination du couple d'agents



négociants);

- **PROP** : utilisé par l'agent initiateur pour faire une proposition à l'agent participant. Ce type est utilisé uniquement dans l'état S6 (Etat de négociation);
- **CPRP** : utilisé par l'agent participant pour faire une contre proposition à l'agent initiateur. Ce type est utilisé aussi dans l'état S6 ;
- **ACPT** : utilisé par l'agent participant (resp. initiateur) pour accepter la proposition (resp. contre-proposition) de l'agent initiateur (resp. participant). Ce type de message est utilisé uniquement dans l'état S6 ;
- **CRNT** : utilisé pour actualiser les valeurs des incohérences après ré-ordonnancement local. Ce type de message est utilisé par les agents en attente dans l'état S5 (Etat d'attente);
- **FNGC** : ce type exprime la fin d'une négociation, et permet aux agents en état d'attente (état S5) et aux agents négociants de passer à une nouvelle étape de détermination du couple ayant la plus grande incohérence (état S4) afin d'entamer une nouvelle négociation ou bien à l'état final (état SF) dans le cas où toutes les incohérences ont été éliminées.

Le pseudo-code 1 illustre un exemple sur la gestion de la file d'attente des messages reçus durant l'état S5:

Pseudo-code 1 : Gestion de la file d'attente dans l'état S5

```

1 :Do
2 : message = receive();
3 :if message.type ≠ FNGC and message ≠ CRNT then
4 : Mettre le message à la fin de la file d'attente
5 :end if
6 :while (message.type ≠ FNGC and message ≠ CRNT) ;

```

7 : Traiter le message ;

Les messages qui seront traités dans cet exemple sont seulement les messages de fin de négociation **FNGC** (qui

permettent au passage à prochaine négociation) et les messages de coordination **CRNT** (pour une mise à jour des connaissances après un ordonnancement local effectué par l'un des agents négociants). En effet, après la lecture du message depuis la file d'attente (Ligne 2), l'agent procède à une vérification de son type (Ligne 3). Si ce dernier ne correspond pas aux types FNGC et CRNT alors l'agent le remet à la fin de la file d'attente (Ligne 4) et passe au message suivant, dans le cas contraire, l'agent procède à un traitement qui dépend du type de message lu (Ligne 7).

V. CONCLUSION ET PERSPECTIVES

Dans ce papier, nous avons présenté une approche multi-agents pour la résolution distribuée du problème d'ordonnancement des ateliers job-shop. La démarche de résolution exploite au mieux la nature distribuée de la fonction ordonnancement en respectant l'autonomie des différents acteurs et en utilisant le paradigme des systèmes multi-agents. En effet, chaque machine est assimilée à un agent autonome qui gère son propre ordonnancement, et qui coopère avec les autres agents à travers un protocole de négociation de manière à assurer une cohérence entre les décisions prises localement. Pour valider l'approche proposée, nous avons opté pour l'utilisation de la plateforme JADE afin de développer le système d'ordonnancement multi-agents. Une modélisation du comportement des agents sous forme d'une machine à état finis est proposée. Cette dernière permet aux agents d'adopter les actions appropriées selon la situation à laquelle ils sont confrontés. Nous avons aussi défini plusieurs types de messages servant d'échange lors des communications entre les agents ainsi qu'une règle de gestion des files d'attente des messages reçus au niveau d'un agent afin de résoudre les problèmes liés à la synchronisation et cela afin d'assurer un fonctionnement global correct du système. Actuellement, nous sommes en phase d'implémentation de l'approche pour sa validation à travers des tests expérimentaux sur des benchmarks tirés de la littérature.

REFERENCES

- [1] Gotha, "Les problèmes d'ordonnancement", *RAIRO Recherche Opérationnelle/Operations Research*, vol. 27, no.1, pp. 77-150, 1993.
- [2] A.J. Davenport & J.C. Beck, "A survey of techniques for scheduling with uncertainty", <http://www.eil.utoronto.ca/chris/chris.papers.html>, 2000.
- [3] W. Herroelen & R. Leus, "Robust and reactive project scheduling: a review and classification of procedures", *International Journal of Production Research*, vol. 42, no. 8, pp 1599-1620, 2004.
- [4] J. Ferber, "Les systèmes multi-agents: Vers une intelligence collective", InterEdition, 1995.
- [5] M. J. Shaw and A. B. Whinston, "Distributed planning in cellular flexible manufacturing systems", *Manage. Inform. Res. Center, Purdue Univ., West Lafayette, IN, Tech. Rep.*, 1983.
- [6] A. D. Baker, "Manufacturing control with a market-driven contract net", Ph.D. thesis, Rensselaer Polytechnic Inst., Troy, NY, 1991.
- [7] J. Liu and K. P. Sycara, "Distributed problem solving through coordination in a society of agents", presented at the 13th Int. Workshop on DAI, 1994.
- [8] A. Agnetis, P.B. Mirchandani, D. Pacciarelli, A. Pacifici, "Nondominated schedules for a job-shop with two competing agents", *Computational and Mathematical Organization Theory*, 6(2), pp. 191-217, 2000.
- [9] A. Agnetis, P.B. Mirchandani, D. Pacciarelli, A. Pacifici, "Scheduling problems with two competing agents", *Operations Research*, vol. 52, no.2, pp. 229-242, 2004.

- [10] S. Ourari, "De l'ordonnancement déterministe à L'ordonnancement distribué sous incertitudes", Thèse de Doctorat soutenue à l'Université Paul Sabatier de Toulouse, France, 2011.
- [11] A. Kouider, S. Ourari, B. Bouzouia, M. Mihoubi, "Approche Multi-Agents pour l'ordonnancement dynamique d'atelier de production", 9ème Conférence Internationale de MODélisation et SIMulation, MOSIM'12, 06 au 08 Juin, Bordeaux, France, 2012.
- [12] F. Bellifemine, A. Poggi and G. Rimassa, "Developing Multi-agent Systems with JADE", published in Intelligent Agents VII, LNAI 1986, pp. 89–103, 2001.
- [13] Foundation for Intelligent Physical Agents, "FIPA 97 Specification. Part 2, Agent Communication Language", 1997.
- [14] J.K. Lenstra, A.H.G. Rinnooy Kan & P. Brucker, "Complexity of machine scheduling problems", Annals of Discrete Mathematics, vol. 1, pp. 343-362, 1977.
- [15] J. Erschler, G. Fontan, C. Merce & F. Roubellat, "A new dominance concept in scheduling n jobs on a single machine with ready times and due dates", Journal of Operation Research, vol.1, pp. 114–127, 1983.
- [16] J.Adams, E.BAlas & D.Zawack, "The shifting bottleneck procedure for job shop scheduling", Management Science, vol. 34, n. 3, pp. 391-401, 1988.
- [17] L. Trung, "Utilisation d'ordre partiel pour la caractérisation de solutions robustes en ordonnancement", Thèse de doctorat, Laboratoire d'analyse et d'architecture des systèmes (LAAS),CNRS, Toulouse, France, 2005.
- [18] A. Singh, D. Juneja, A.K. Sharma, "Agent Development Toolkits", International Journal of Advancements in Technology, vol. 2, no. 1, pp. 158-164, January 2011.

Coordination based agents for pathfinding

Mohammed Chennoufi

*Computer mathematics faculty,
University of Science and Technology
of Oran BP1505, Oran 31000
Algeria*
Mohammed.chennoufi@univ-
usto.dz

Fatima Bendella

*Computer mathematics faculty,
University of Science and Technology
of Oran BP1505, Oran 31000
Algeria.*
Fatima.bendella@univ-usto.dz

Maroua Bouzid

*University of Caen, Basse-
Normandie Campus Cote de Nacre,
BP 5186 - 14032 Caen France*
bouzid@unicaen.fr

Abstract— This paper describes a new algorithm for decentralization of the A* algorithm called DEPA*, which is a quick algorithm for finding the shortest path between two nodes in a graph. This algorithm is based on an heuristic evaluation to each vertex to estimate the best path that passes until you reach the goal. For this, we propose an approach based on multi-agent systems that decomposes the graph into sub related graphs (many agents). This connection is guaranteed thanks to the characteristics of agents all computing an A* at each sub-graph in a parallel way. The initial and final state of each agent will be chosen according to a well defined heuristics. A coordinator agent resolves the conflict in the case of many final states in a sub graph. Then the agents interact to achieve the goal. We illustrate this approach on a grid connected by square cells.

Keywords—Agent; A*; DEPA*; Grid; Node; Multi-Agent Systems; Dijkstra.

I. INTRODUCTION

The distributed approach is presented in our daily lives such as traffic for example to go from city A to B can escalate the City B, same for the Web Service, if you ask query <how to get from A to B>, the answer may be passed through several server, planning in a travel agency is also regarded as a distributed approach (car, bus, avian ...).

As the world develops, the need to move has become a fundamental element, which makes systems more complex. Recently there has been a growing interest in multi-agent planning [1][2]. The problem is to compute a path for each agent from an initial location to a goal without have conflict with other agents. However, to our knowledge, very few researchers have applied the software agents on the shortest path problems.

Our objective is to design an intelligent and effective DEPA* able to accelerate the search of the shortest path between two nodes if it exists. Our algorithm is developed to find an optimal and complete solution. We illustrate our approach on a grid, this grid is divided into several agents: Some A* are run in parallel in accordance with the concept of depth (A* that do not meet certain conditions are not executed), each Agent calculate the shortest path as well as the successor to its final state. A coordinator agent solves the problem of conflict if it exists by calculating the minimum distance to the final state in the sub graph or global heuristics with adjacent agent.

This article is organized according to the following provisions: In the first section, we present previous works in the field of distributed planning and the variations of A*. In the second section, we discuss on the modeling of the problem and the explanation of A*. The third section is

devoted to the presentation of our approach based on the agents with a pseudo code. Results are given in the fourth section. Conclusion and perspectives are presented in the last section.

II. RELATED WORKS

The Dijkstra algorithm (Research in width) [3] and its variations as A* so-called A START [4] are well known algorithms in artificial intelligence, specifically in planning (the calculation of shortest path).

Our state of the art is divided into three major axes. The first is on the extensions of A*, then on the planning distributed and finally on decentralization of A*. [5] Proposes an algorithm of shortest path based on A* algorithm, its main advantage is in the use of a bounded memory while the A* algorithm uses exponential memory. All other features are inherited from A*. It avoids repeated states as long as the related memory permits it. In [6], the author introduces the notion of time, in order to accelerate repeated heuristic A* research with the same state toward a goal. The idea is to place dates at the state level in its local search spaces, it makes the heuristics better informed after each A* heuristic search. The work of [7] are inspired from [8] Dynamic A* which behaves like A*, except that the costs of arc can vary as the algorithm works. Two other states are added, it is Raise which indicates that its cost is higher than the last time it was on the Open list, and lower indicating that the cost is less than the last time it was on the Open list. [9] Proposes an incremental version of A* so-called D*. The idea is to take advantage of previous researches that they reuse after repair. These changes allow the graph of a well gained execution time better than rescheduling from zero. In [10], the authors combine both incremental and heuristic searches. They reuse information from previous searches to speed up searches for similar sequences. GAA* (Generalized Adaptive A*) solves search problems potentially faster. The heuristic search often based on A* heuristics uses heuristics knowledge in the form of approximations of goal distances. GAA* is much faster than uninformed search algorithms. Recently, another algorithm [11] has been proposed. It is modeled as a grid it depends on a forward chaining search called FRA* (Efficient Incremental Search for Moving Target Search). Whenever the target moves, FRA* quickly adapts the search tree previously built to the new target position and recalls the function of A* search. This adaptation of the tree is largely dependent on environmental modeling. The changes in the environment are not treated. [12] Developed a variant of this algorithm called GFRA* (Generalized Fringe-Retrieving A*) which enables FRA* to operate in environments modeled by arbitrary graph, in addition to the heuristic function used is ineligible. This is the case in the field of planning. In [13]

the author resolves collisions; he added the concept of priority in the context of cooperation between agents in finding the shortest path. Each agent is assigned to a priority where agents will be executed in the order of this priority. In [14] the authors present an asynchronous variant of decentralized planning, exploiting the parallelism in distributed systems, which gives a speed in calculations. Unlike synchronized planning approaches, the algorithm allows an agent to call his local planner of the spatial-temporal trajectory to find the best path. [15] Propose a complete algorithm that is fast enough for real-time applications based on MAS, at any time when an agent finds a solution. It uses the rest of the time to gradually improve the solution until it is optimal. The algorithm can solve problems at the rate of 80% with small grid but with the scale, the problem remains. In [16] the cooperation between agents illustrated by the author is based on two algorithms: First, to find the interface agents that should cooperate to answer a query "how to get from A to B?", and how to make these agents cooperate. The two algorithms are based on the Dijkstra where each agent details the itinerary. The author did not unfortunately give a pseudo code that clarifies how this distribution is achieved.

Work on the decentralization of A* is little although the areas covered are decentralized in nature, such as: road traffic, web service, game theory [17] Proposed a distributed search algorithm DEC-A* (a Decentralized Multiagent Pathfinding) modeled as follows: when a problem is presented, each agent calculates its overall heuristic that estimates the cost of its shortest path to the goal intermediary to its neighboring agents. Then the agent containing the initial state develops locally in A* by minimizing the cost until it reaches the border then it stimulates on the other side a new execution of A*. These steps are repeated to reach the goal. So, the author has made an extension of the heuristic evaluation of the distance like the sum of two functions, a local, which evaluates the cost to reach the closest node, neighbor to the objective and the overall distance that estimates the cost of sub graph in the target through other graphics. His work decreases the time to find the shortest path and reduces the complexity but it is not effective in a graph without hindrance.

In this domain, several studies have been carried out especially in the last 10 years, but we noticed that most of the researchers work on variations or extensions of A* applied in different areas such as game theory, robotics, traffic road.... But little work on decentralization of A* especially on coordinating agents at boundaries nodes, that is the object of our work

III. MODELLING OF THE PROBLEM

A. Notation

Our approach is illustrated in a square grid that represents blocked cells in grey and unblocked in blond. The initial state is colored in green and red for the end state. S is the finite set of unblocked states, $s_{init} \in S$ is the initial state (starting node) of our research, $s_{goal} \in S$ represents the final state (arrival node), $h(s) \in S$ (heuristics) is the approximation of a current state s to a state s' . In our illustration we work with the Manhattan distance (Pearl J., 1985).

$h(s) = |x_s - x_{s'}| + |y_s - y_{s'}|$, where x and y are the coordinates of the cell. $C(s, s') = 1$ is the cost of transition between

adjacent cells in the 4 directions $\rightarrow \leftarrow \downarrow \uparrow$, $C(s, s') > 0$ is the cost of transition between the cell $s \in S$ and $s' \in S$. $G = \{G_1, G_2, G_3 \dots G_n\}$ where G_n is the n th sub graph, each G_i is a sub graph of agent A_i , the agents communicate using a border node $Succ(S_{ij}) = S'_{ij}$, $s \in G_i$ and $s' \in G_j$ with $i \neq j$, let $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ a set of path to each sub graph starting from an initial border state of G_i to a final boundary state G_i .

A* use the formula $F(x) = g(x) + h(x)$ which is the current approximation of the shortest path to the goal, where $g(x)$ is the total distance between the initial position to the current position and $h(x)$ is the heuristic function which represents the approximate distance from the current location to the goal state. $g(As_{init})$ is the minimum cost for the A_i agent and $h(As_{init})$ is the minimum heuristic of the same agent.

B. Standard A* algorithm search

A* pronounced A Star [4] is a algorithm search of artificial intelligence that performs a heuristic search [18] in an area to find an optimal path from the root node to the goal node. The algorithm search is based on a heuristic evaluation between two nodes in order to eliminate many paths of high costs, Two representations are possible tree or grid in Fig. 1 and Fig. 2.

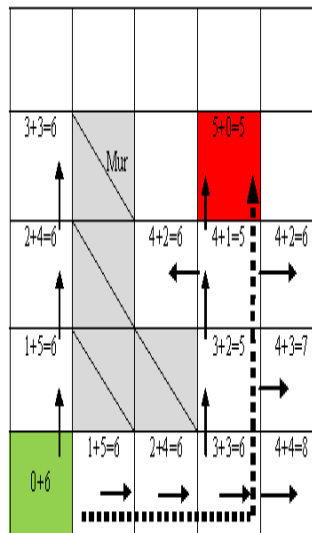
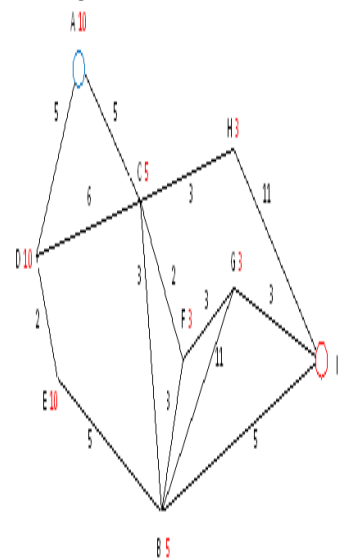


Fig. 1. Representation in greed and tree

To find a path from one point to another we must begin by heading to the destination. It is precisely this idea that the A* algorithm uses. The idea is very simple: at each iteration A* will try to get closer to the destination, it will therefore focus on possibilities directly closer to the destination, putting aside all others. All possibilities not allowing to get closer to the destination are set aside, but not suppressed. They are simply put in a list of possibilities to explore (open list) if ever the solution currently explored is poor. Indeed, we cannot know in advance whether a path will lead or be shorter. Enough for this path lead to a dead end that the solution becomes unusable. The algorithm will therefore first move towards more direct paths. And if these paths fail or prove wrong later, it will examine solutions being put aside. This going back to review the solution that was set aside the algorithm ensures that we always find a solution (if it exists, of course). What makes this algorithm search complete, fast and optimal Fig. 2. Several distances can be used like the Manhattan distance, diagonal, Euclidean...

An illustrative example of A* running on a grid and a tree with the trace of open list and closed list is shown in Fig. 2

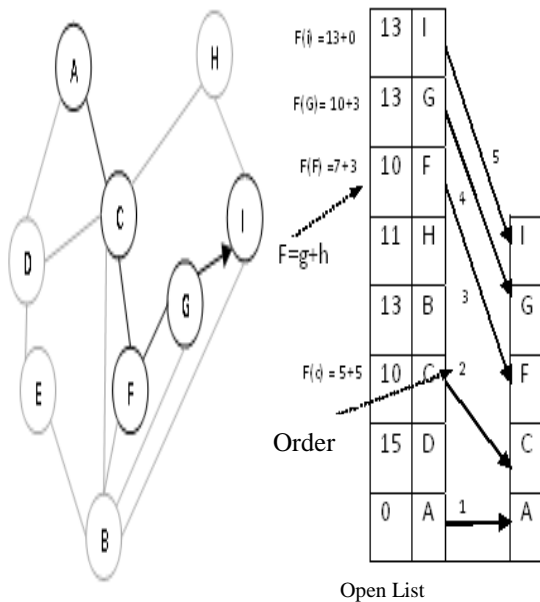


Fig. 2. Illustrative example of A*

IV. PROPOSED APPROACH

Our idea is simple but effective. After the decomposition of the graph in sub graph (sub-matrix). Each agent calculates respectively the initial state which represents the minimum cost of the border states and the final state representing the minimum of the heuristics of the border states. Then A* will execute in parallel with the sub graphs provided that its initial state is a successor of the final status of another sub graph (red crosses in Fig. 3 for the case of untreated parallel). In the case of a single processor a multi thread is run. The agents communicate using an coordinator agent which calculates

the overall heuristic [17] of each agent. It will be used in case of conflicts between the final states of each sub graph and build the final path from S_{init} to S_{goal} .

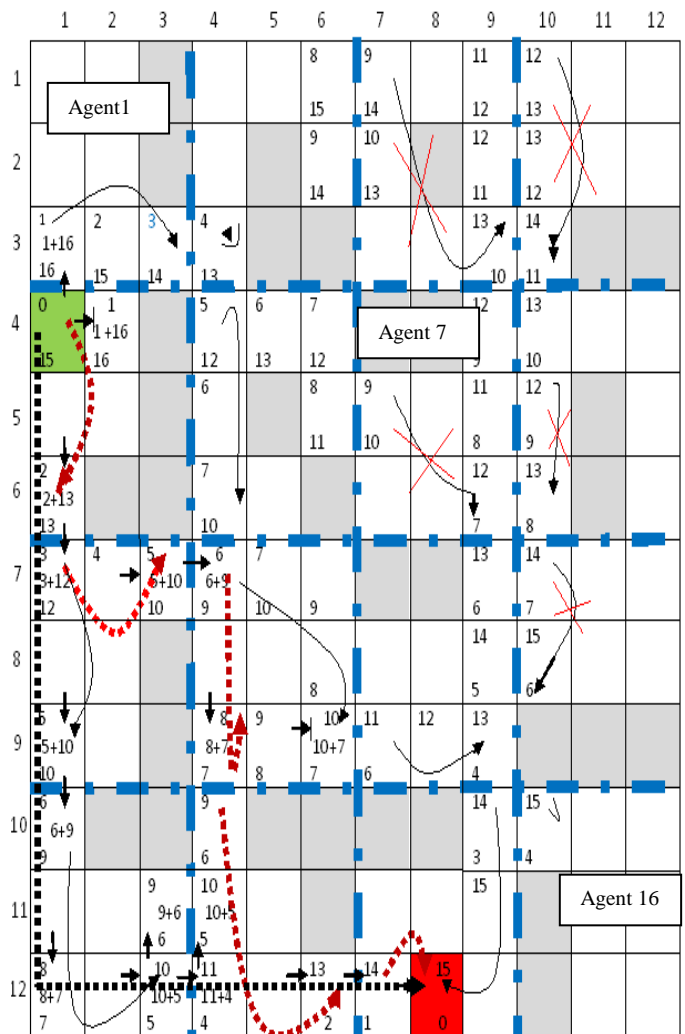


Fig. 3. illustration of an example DEPA*

Here is a sample execution which explains the algorithm 1 and 2: Our grid is divided into 16 agents A_{ij} from A_1 to A_{16} . In each square the value presented at the upper left corner represents the cost. The bottom left represents the heuristic and $f = g + h$ in the middle of the cell. All initial and final states are stored in two vectors provided that the initial state is a successor of the final state.

The A_9 agent its initial state is $s_{7,1}$ which is a successor of $s_{6,1}$ for A_5 agent. So A_9 will be considered contrary to the A_7 agent as its initial state $s_{5,7}$ is not a successor of $s_{5,6}$ of A_6 agent. Then we calculate the A* for agents well sorted and we appeal to the coordinate agent which regulates the conflict in case there are more final states in the same agent (two A* in the case of A_9 and A_{10} agent). The Fig. 4 shows an example of cooperation between agents where each agent sends a message send () and receive message answer () of the coordination agent which contains the partial path λ_i

4.1 Pseudo code

Algorithm 1: DEPA*

```

1:Begin
# Initialization
2:Decomposition (grille, nbr noeud);
3:For each  $A_i$  of Ado
# Calculate the initial and final state of each sub graph
4:  $s_{A_{init}} = MIN(g(s_{ij})$  of node neighbor in  $A_i$ );
5:  $s_{A_{final}} = MIN(h(s_{ij})$  of node neighbor in  $A_i$ );
# Condition for parallelism
6: If  $s_{A_{init}} == Succ(s_{A_{j_{final}}})$   $i \neq j$  then
7:  $Vector1 = s_{A_{init}}$ ;
8:  $Vector2 = s_{A_{final}}$ ;
9:End of if
10: End of for
11:For  $i=1$  to  $length[Vector1]$  do
# calculate of A* for each agent in parallel
12:Thread[ $i$ ].  $A^*(Vector1[s_{A_{init}}], Vector2[s_{A_{final}}])$ ;
13: End of for
# Communication between agents
14:Cordiagent ( $A_i$ );
15:Pathfinal =  $\lambda = \sum_1^n best\lambda_i$ ;
16:End.
    
```

The line 2 of the algorithm 1 represents the decomposition of the grid in sub grid depending on the desired number of agents. Take our example for a grid of 12*12. If we want have agents with 09 nodes. We must split the rows and columns in 3.3 which give us 16 agents (4*4). The 4 and 5 lines calculate the minimum of the costs and heuristics for boundary nodes for each agent. They will be stored in two vectors (lines 7 and 8). Much implementations is shown in line 12, which uses the standard function A* in parallel. Line 14 calls he coordinator agents (algorithm 2)

The Algorithm 2 guaranteed the communication between Agents (boundary nodes). It used to calculate the best paths for

each agent (line 12). Three tests are performed (lignes 3, 6.9) to solve the problem of multiple final states. Line 10 calls the heuristic function $h_{global}(s_{A_{ij}})$, which represents the minimum distance for A_i to the final state s_{final} . is calculated by a chainigback for $s_{final} = s_{12,8}$ to the initial state $s_{init} = s_{4,1}$ [17].

DEPA* is complete, unless there is an infinite number of nodes with $f \leq f(G)$. Since h is admissible, $best\lambda_i$ is optimal because it is a simple A*. So $\sum_1^n best\lambda_i$ is optimal.

Algorithm 2: Cordiagent (A_i)

```

1:Begin
# Initialization;
2: For each node of  $A_i$  do
3:If  $h(s_{A_{final}}) == h(s'_{A_{final}})$  then
4:  $s'_{final_{ij}} = MIN(c(s_{A_{init}}, s'_{A_{h_{final}}}) / s' \in A_i)$ 
5:End of if
6:If  $c(s_{A_{init}}, s'_{A_{h_{final}}}) == c(s_{A_{init}}, s''_{A_{h_{final}}})$  then
7:  $s'_{final_{ij}} = MIN(Succ(h(s_{A_{ij}})), Succ(h(s'_{A_{ij}})))$ ;
#  $s' \in A_j$   $j \neq i$ 
8:End of if
9:If  $h(succ(s_{ij})) == h(succ(s'_{ij}))$  then
10:  $s'_{final_{ij}} = MIN(Succ(h_{global}(s_{A_{ij}})), Succ(h_{global}(s'_{A_{ij}})))$ ; #  $s' \in A_j$   $j \neq i$ 
11:End of if
12:Return  $best\lambda_i$ ;
13:End of for
14:End.
    
```

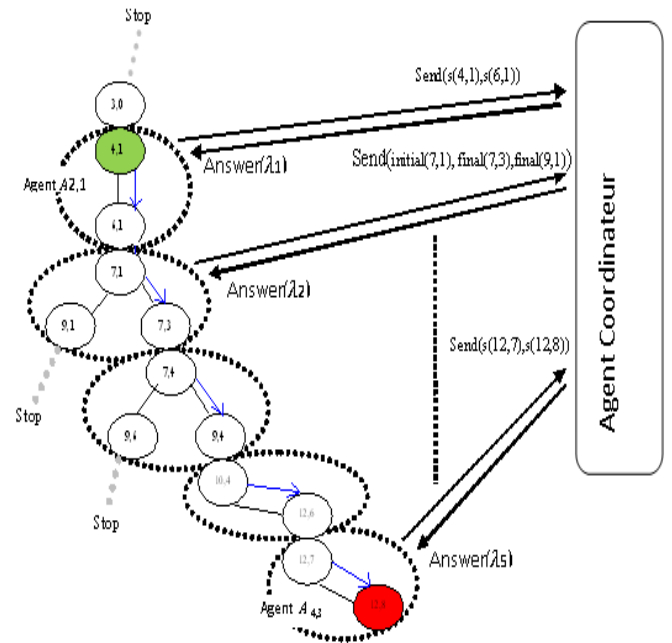


Fig.4. Coordination between agents

V. EXPERIMENTATION

We have compared our DEPA* algorithm with A* and DECA* on a machine with an Intel (R) Core 2 Duo 3.16 GHz CPU and 2 GB of RAM. After several experiments, we obtained results that illustrate the execution time with respect to the number of input nodes and the number of agents.

Obstacles in our grid that represent walls, rivers, mountains are programmed in a way random all depends on the problem. We used a Boolean function. When creating the grid. It is true for adjacent nodes (not obstacle) if Math

.random exceeds 0.1.

TABLE I. EXPERIENCE WITH DIFFERENT GRID

Grid size	A* en ms	DECA* en ms	DEPA* en ms
12*12	32	15	15
20*20	47	17	16
50*50	218	256	150
100*100	1000	410	300

The first aspect we noted in Fig.5 is the execution time of DEPA* which calculates the shortest path from node (0, 3) to (40, 40). This time with DEPA* is the least by contribution to A* and DECA*. Even if the grid is large (table 1: 10000, 25000 nodes).

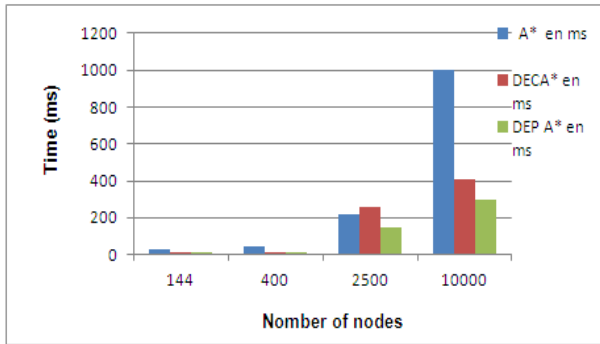


Fig.5. Comparison between variations of A*

TABLE II. EXPERIENCE WITH A GRID (30*30)

Grid size	Number of agents	A* ms	DECA* ms	DEPA* ms
30*30=900	10*10 agents of 3*3 nodes	153	145	16
	2*2 agents of 15*15 nodes	110	59	38

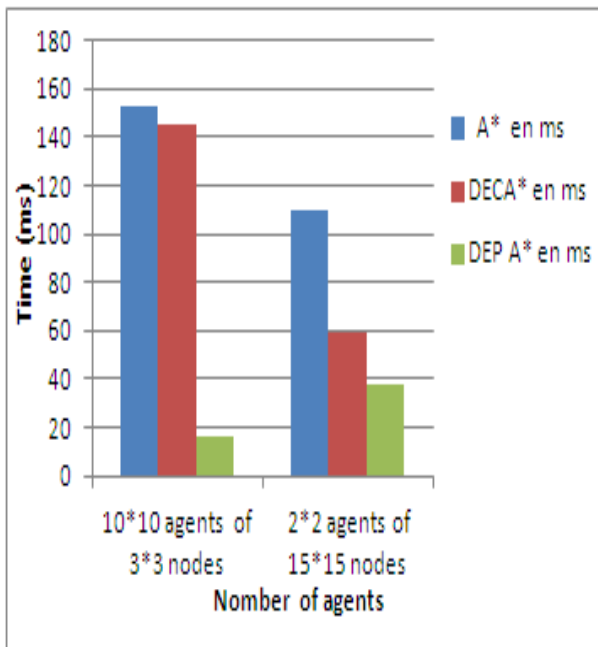


Fig. 6. Grid of 30*30 nodes

In Fig. 6 and Fig.7, which illustrate the tables 2 and 3, we notice the profit in the execution time in relation with the number of agents for 03 algorithm. DEPA* is the best in the time of executions. Consider the case of Fig. 6, which shows a grid of 30 * 30 with initial and final states are (0,3) and (29,29), for a breakdown of 10 * 10 agents where each agent has 3 * 3 node. DEPA* runs in 16 milliseconds but in the decomposition of 2* 2 agents with 15 * 15 node, DEPA* runs in 38 ms which expresses the power of parallelism and coordination between agents.

The Fig.8 shows the performance of the DEPA* on scale with initial and final states are at the end of the grid: (3,3) and (170,170). We notice that as the number of agents increases the time is decreased 983 ms for 4 agents and for 36 agents we have 109ms which expresses a time saving through the parallel execution of A* at each agent.

TABLE III: EXPERIENCE WITH A GRID (50*50)

Grid size	Number of agents	A* ms	DECA* ms	DEPA* ms
50*50=2500	10*10 agents of 5*5 nodes	218	57	15
	2*2 agents of 25*25 nodes	313	413	94

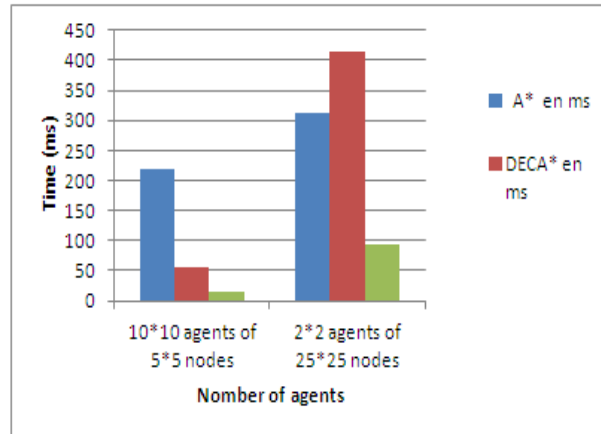


Fig. 7. Grid of 50*50 nodes

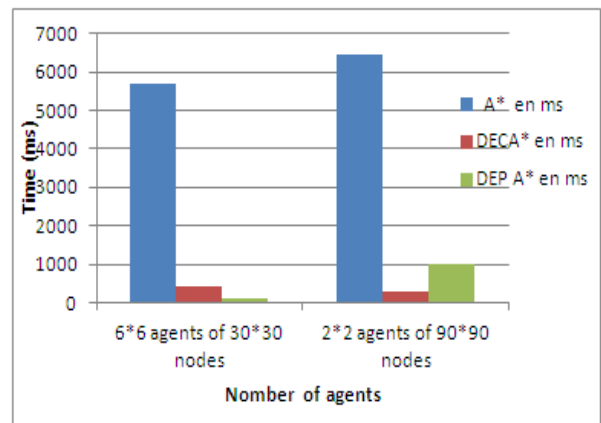


Fig. 8. Grid of 180*180 nodes

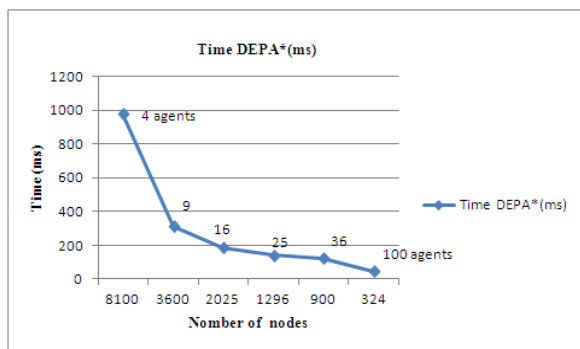


Fig. 9. Grid of 180*180 nodes

The Fig. 9 shows the relationship between the increase in the number of agents contribution to the reduction of time execution and the number of nodes for each agent on a grid of 180 * 180 nodes using the parallelism of A*.

We notice that, in our experiments, DEPA* finds a solution if it exists in at least a second with grids that can contain up to 32400 nodes.

VI. CONCLUSION AND FUTURE WORKS

Networks grow continuously, which makes the system more complex. More recent works in artificial intelligence handle the problem of shortest path. The multi-agent systems are helping to solve this complexity with a decentralized manner through a "send and answer" communication and the coordination between agents in order to achieve the goal.

The proposed DEPA* algorithm is in keeping with this problem. We have illustrate our approach on a square grid like the game grid, which allows to compute the shortest path from an initial state of an agent to a final state of another. It is based on the parallel A* that is run on agents of which their initial states are successors of the final states of other assistants agents. Much of the work is devoted to the coordination between agents to arrive at the final path in the case of several A* (several final states) in the sub grid (agent). The coordinator agent regulates this conflict: it takes the final state which corresponds to the minimum cost in relation to the initial state of the same agent. In the worst case the coordinator agent consults the global heuristic [17] of the assistant agents. The final state chosen is the one that corresponds to a minimum of global heuristic.

We obtained good results, especially in a scale of 25000 nodes and 32400 nodes where DEPA* finds the solution if it exists in some milliseconds. In the future, and for the

future researches, we are interested in the extension of the field of application, such as the simple and dynamic graphs, when the final state changes in position.

REFERENCES

- [1] M. Erdmann, and T. Lozano-Perez, "On multiple moving objects," *Algorithmica* 2:477-521. 1987.
- [2] K. Fujimura, "Motion planning in Dynamic Environments," New York, NY: Springer-Verlag 1991.
- [3] PE. Hart, NJ. Nilsson and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems, Science, and Cybernetics*, 100-107, 1968.
- [4] S. Russell, "Efficient memory-bounded search methods," In Neumann, B. Proceedings of the 10th European Conference on Artificial intelligence. Vienna, Austria: John Wiley & Sons, New York, NY. pp. 1-5, 1992.
- [5] S. Koenig and M. Likhachev, "AAMAS," Hakodate, Japan May, 2006.
- [6] S. Koenig, M. Likhachev and D. Furcy, "Lifelong Planning A*," *Artificial Intelligence Journal*, 155, (1-2), 93-146, 2004.
- [7] S. Koenig and M. Likhachev, "Adaptive A*," In Proceedings of the International Conference on Autonomous Agent and Multi-Agent Systems, p. 1311-1312, 2005.
- [8] S. Koenig and M. Likhachev, "D* lite". In Proceedings of the Association for the Advancement of Artificial Intelligence, p. 476-483, 2002.
- [9] S. Koenig, M. Likhachev, Y. Liu and D. Furcy. "Incremental Heuristic Search in Artificial Intelligence," *Artificial Intelligence Magazine*, 25(2), 99-112, 2004.
- [10] X. Sun, S. Koenig and W. Yeoh. "Generalized adaptive A*," In Proceedings of the International Conference on Autonomous Agents and Multiagent Systems, p. 469-476, 2008.
- [11] X. Sun, W. Yeoh and S. Koenig, "Efficient incremental search for moving target search". In Proceedings of the International Joint Conference on Artificial Intelligence, p. 615-620, 2009.
- [12] X. Sun, W. Yeoh and S. Koenig, "Generalized Fringe-Retrieving A* : Faster Moving Target Search on State Lattices". In Proceedings of the International Joint Conference on Autonomous Agents and MultiAgent Systems, volume 1081-1088, 2010.
- [13] S. David, "Cooperative Pathfinding," Proceedings of the First Conference on Artificial Intelligence and Interactive Digital Entertainment: pp. 117-122, 2005.
- [14] M. Cáp, P. Novák, J. Vokřínek and M. Pechoucek, "Asynchronous Decentralized Algorithm for Space-Time. Cooperative Pathfinding," Workshop proceedings ECAI 2012.
- [15] T. Standley and R. Korf, "Complete Algorithms for Cooperative Pathfinding Problems," Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 2011.
- [16] M. Kamoun, "Designing an information system using multimodal travel : A multi-agent approach to search and composition routes online", PHD thesis Lille university, 2007.
- [17] M. EL Falou, M. Bouzid and M. Mouaddib, "DECA*: a Decentralized Multiagent Pathfinding Algorithm", AAAI 2012.
- [18] J. Pearl. "Heuristics: Intelligent Search Strategies for Computer Problem Solving," Addison-Wesley, 1985.

Replication Strategies in a Mobile Environment

Moufida Rahmani

LSI, Computer Science Department
University of Science and Technology Houari
Boumediene Algiers, Algeria
morahmani@usthb.dz

Mahfoud Benchaïba

LSI, Computer Science Department
University of Science and Technology Houari
Boumediene Algiers, Algeria
mbenchaiba@usthb.dz

Abstract—The applications in mobile ad-hoc network (MANET) like peer to peer (P2P) file sharing have attracted more and more attention in recent years. However, due the frequent network partitions caused by node mobility and the limited energy, the data availability is very low. The replication can be used as the efficient technique to improve resource availability, enhance the system performance and achieve load balancing. This paper surveys some existing data replication protocols that have been proposed in the literature for MANET and Mobile P2P networks and compares them.

Keywords—P2P; MANET; MP2P; Replication; Availability; Access frequency.

I. INTRODUCTION

A mobile ad-hoc network (MANET) is a set of autonomous mobile nodes (e.g., laptops, PDAs, mobile phones) that form a temporary network with variable topology and communicate using wireless links. Several applications have been deployed in a MANET to provide data services.

Peer to Peer (P2P) is an overlay network that allows direct sharing of resources such as CPU, bandwidth, files etc, among a large number of users in a decentralized manner. It is nowadays widely adopted on the Internet and is used for many different application types, such as file sharing (Gnutella [1] and KaZaA [2]), communication as Skype [3] and distributed computing as SETI@home [4].

MANET and P2P networks are both decentralized and self-organizing networks with dynamic topologies. This common features have attracted a lot of attention in the research community to deploy P2P applications for MANET (MP2P). However, these systems face many constraints such as the limited physical resources like bandwidth and energy and node mobility. This leads to frequent network partitions or performance degradation which makes data available in MP2P networks slow. Data replication is a feasible solution to increase data availability as well as enhance the system performance by reducing communication overhead, reduce the data querying delay and achieve load balancing.

To the best of our knowledge, Cohen et al [5] were the first authors that addressed strategies of replication in unstructured P2P networks. The goal of their work is to replicate in order to reduce random search times, i.e. to minimize the number of peers that have to be probed before the query is resolved, therefore the system performance is improved. Many other replication methods are proposed for the structured and unstructured P2P networks [6][7], but these methods are proposed for the fixed network (Internet) where the mobility of the node was not considered. Therefore, they cannot be implemented in a mobile environment. In the case of MANET, Hara [8] was the first that introduces data replication problem. He proposed three methods to assign the replicas to mobile nodes in order to improve the availability of shared data.

In this paper, the design issues concerning developing a replication protocol for MANET have been discussed. We describe various of the existing replication techniques for MANET and MP2P networks and compare them. The aim of this paper is to provide a comprehensive study that helps the designer to choose the most appropriate data replication strategies for MANET.

The remainder of this paper is organized as follows. Section II discusses the data replication in a mobile environment. Section III reviews the replication techniques proposed for MANET and section IV reviews the replication techniques for MP2P. Section V presents a summary of all the replication techniques discussed in this paper. Finally, Section VI concludes the paper.

II. REPLICATION IN MOBILE ENVIRONMENT

Data Replication is the process of copying the data from a source node to other nodes. It is an indispensable solution for MANET to enhance data availability as well as reduce communication overhead, achieve fault-tolerance and load balancing for the applications deployed on top of it. As replication has advantages, it also has significant costs such as the storage cost and consumption of bandwidth. Therefore, the excessive use of replication is not recommended i.e. the mobile node cannot have replicas of all data items in the network. Thereby, in order to avoid wastage of network and node resources, the replication technique or called also a replica allocation technique must take into consideration important factors such as what data to be replicated, where the replica to be placed and when starts replication process, etc... In [7], Rahmani and Benchaïba present a study on replication in P2P systems, including: factors involved in replication, type of replication, and parameters affecting replication. Replication can be executed in parallel with lookup protocol, i.e. every time the node receives requests for a data, it launches replication process, if it is necessary. For example, if access frequencies for the required data exceeds a certain threshold. It can also be executed periodically.

A. Classification of replication technique

Many replication techniques are proposed to increase data availability in a mobile environment. These techniques are different to each other, but we can classify them into two categories according to the level of autonomy of the nodes: group-based and fully decentralized strategies.

Group-based strategies: In this category, the nodes are organized into different groups. Each group is managed by a clusterhead (CH) or super peer (SP) which has better performance compared to the other nodes like high energy, high available bandwidth and high processing capacity. CH collects the necessary information such as data items' access frequencies and remaining storage space from the nodes that exist in its group. After that, CH performs replication in its group. This category can

take better decision of replication compared to the fully decentralized strategy because it has some global view of parameters included in replication, e.g., free space storage in a group.

Fully decentralized strategies: Each mobile node plays the same role and takes decision of replication locally based on local and some time on past information. This category is suitable in a dynamic environment with autonomous nodes which have links between them not reliable.

B. Issues related to data replication

Due to the different constraints imposed by such environment: limited energy, low bandwidth, unpredictable node mobility, etc. Data replication in MP2P is very complicated and must deal with the different following issues.

1) Power consumption

The mobile nodes, e.g., laptops, PDAs, mobile phones are fed by energy sources autonomous (batteries), which limits services and applications that can be supported by each node. If popular data items are replicated on the node with low power, then the request for these data items will drain its power. Therefore, a replication algorithm must try to reduce and balance power consumption among the node.

2) Node mobility

The mobile nodes are free to move randomly. Consequently, the network topology may change rapidly and in an unpredictable way. The nodes might move to a place where they cannot be reached by other nodes. Thus, the replication technique should avoid to replicate popular data in such nodes and tries to copy in advance their data (replicas or original) in some other nodes to improve data availability. Therefore, the replication algorithm should predicate or calculate node mobility in advance in order to avoid these problems.

3) Resource availability

Some nodes are small devices, then processing power, memory and storage capacity are limited. The replication algorithm must take into account these limitations, for example the algorithm has to check whether a node has sufficient memory capacity to hold the replica or not, and tries to achieve its goals but with minimum cost.

4) Network Partition

The network partition is caused by the node mobility or limit battery power and it poses a serious problem which is the nodes in one partition are not able to access the data hosted by nodes in the other partition. Therefore, data availability is low and significantly degrades the performance of data access. In order to solve this, the replication technique should be able to determine or predict the time when the network partitioning might occur and replicate data items beforehand.

5) Consistency Management

A request for a data item should always access the most recent copy of the data item. Hence a replication algorithm is required to manage the consistency of data in the network. As the network has poor resources, the global consistency of replicas is not desirable.

III. REPLICATION STRATEGIES FOR MANETS

A. Replica Allocation Technique Based on Clusters (RABC) [9]

In RABC, each mobile node has a unique identifier $node-id$ and maintains a Replica Index RI . RI contains information about the data items, there are 3 entries related to each data item: the id of this data item $item-id$, its size and its Time To Live (TTL) value. The TTL value shows after how much time the data item before to expire. This value is assigned by the owner of the original data item.

Each mobile node is associated with a cluster and each cluster has its cluster head (CH). Every CH maintains an Available Replica Table (ART). ART contains the information about the replicas that are available at different mobile nodes in that cluster. For each node, ART includes node-id, the ids of the data items that are replicated at that node and the free space available at that node.

Replica allocation process (RAP) is started by CH when a search for an object is successful and this CH exists in the path traversed by the response message. In RAP, CH checks its ART, and check for a node with free space available greater than the size of data item. If CH gets a node with space available greater than or equals to the size of data item, then CH replicates the data item on that node. After replicating data item on that node, it sends a message to the CH so that CH will update its entries related to that node in ART. If that node does not exist, then CH chooses a node with maximum free space available. After selecting node, CH redirects that node to apply replica replacement algorithm. In this algorithm, the node will remove data items whose TTL values are expired. After successful creation of free space at that node, CH will replicate that data item on that node, node will send a message to the CH so that CH will update its entries in ART.

When CH receives update message for a data item from the owner of original data item. It broadcasts the updated copy of data item with its item-id to all CHs. Whenever a CH receives an updated copy of the data item with item-id, then CH checks this item-id is in its ART. If desired item-id is found in ART, then updates this data item at all nodes pertaining to item-id in ART. If not exist, then CH starts RAP, so that replica of the data item-id can be created in this cluster.

B. Zone-Based Replication Scheme (ZBR) [10]

The main idea of ZBR is the integration of data replication with data lookup service and underlying zone routing protocol by using the cross-layer design between the routing layer and the application layer. Therefore, it can achieve a flexible and efficient data access service with low overhead.

Each data item has a unique data identifier and its original copy is held by a particular mobile node. Each mobile node has a unique identifier and has certain storage space for creating replicas excluding the space for the original data item.

For each mobile node, a zone of radius r is defined and includes all the nodes whose minimum distance in hops

from the node in question is at most r hops. Each node has 3 tables:

The routing table, is defined in the network layer and maintains routing information to those nodes that are within its zone through some sort of a proactive scheme as OLSR protocol [11].

The data lookup table, is defined in application layer and describes the distribution of the data items (original copies or replicas) within its node's zone. The data lookup table contains the data identifier, a flag that indicates whether the data item is original copy or replicas, a value that indicates the freshness of the data and is used to maintain consistency and a description of the data item.

Capability table provides information about the free storage space of the nodes within its zone.

In order to update data lookup table and capability table with low overhead, ZPR extends hello message sent by OLSR periodically to update the routing table.

When the node wants to access data item, first, it checks its data lookup table. If the data item is available in its local zone, then the process of intra-zone data lookup is invoked as follows: The requester node used its routing table to find the path to the hosting node and to send to it the request. When the hosting node receives the request, it sends the data item to the requester node. If the data item is not available in its local zone, then inter zone data lookup process is invoked as follows: the requester node sends the request to all its peripheral nodes. Peripheral nodes are nodes whose minimum distance to the node is exactly equal to the zone radius. The remaining nodes are categorized as interior nodes. Each peripheral node executes the process of intra-zone data lookup as the same way as the requester node. The same procedure is repeated until the data item is found. When the data item is found (original copies or replicas), the hosting node transfers the data items to the requester node and replication is triggered at the same time.

The replication is triggered when the process of inter-zones is invoked, exactly when the data item is transferred from the hosting node to the requesting node. ZPR replicates only the data item that is not in the nodes within the zone of the requesting node and is replicated into the node with enough free storage space. The replica is first allocated to the peripheral nodes, if they have free storage space, if not, then the replica is allocated to the interior nodes. If the interior nodes do not have enough free storage space, replica replacement procedure is invoked to evict the data replicas. Every data replica is assigned a valid period d , and a data requesting node considers a replica up-to-date if $d > 0$. When replication occurs, the owner decides the validation period d of the data replicas based on updating interval and current data access time if the data requesting node gets the data from the owner. If the data requesting node gets the data from other nodes hosting the replica, the value of d of the new replica equals to the old one. Each node hosting the replica decreases its d value in the same ratio. The replicas with $d = 0$ can be removed from the hosting node to save storage space.

C. Data Replication [12]

The data replication is composed of two main phases: The first phase is executed whenever a new data is created in

the network. The objective of this phase is to inform all the network of the existence of new data and to distribute the replicas uniformly on the network. Therefore, The traffic is reduced and the expenses in energy are balanced. In order to eliminate the redundancy of replicas among the neighbors, this method uses the number of hops to estimate the distance between a node to avoid to create the replica on two nodes neighbors.

The second phase is devoted to the redistribution of replicas in order to overcome the impact of dynamic changes of topology and to satisfy the evolution of users' needs. It is based on the frequencies access data. Each node estimates two types of frequency: The first one is a frequency of external access which represents the access rate of the node to external data which means outside of its cache. If the frequency of external data exceeds a certain threshold, the replication of this data is initiated. This type of replication is called replication on need. The second is frequency of internal access represents the access rate of other nodes to internal data in its local cache. When the node holder of data (original or replica) receives a message of the access demand to this data, it checks if the frequency of internal access of the node requestor for this data extends a certain threshold and the distance between the holder and requestor node is greater or equal to threshold (three number of hops), then instead of sending to it just a response to its demand, it transmits to it data replica. This type of replication is called replication on demand.

D. User-Centric Replication Model (CreaM) [13]

CreaM is a fully decentralized solution. Its idea is to give priority to user needs and allows them to specify their level of participation in the system. Each node makes locally replication decisions, when it is overloaded the replication process is automatically started.

CreaM is composed of several components, the fundamental among them is the Peer State Manager (PSM). The PSM controls the replication process; it monitors the peer's status and decides when to replicate a data item (DI) and when to accept placing replicas on the node. The peer's status is determined by the consumption of its resources. Each user specifies the allowed level of resource consumption, there are three thresholds: the allowed load level of the nodes CPU, the allowed level of remaining battery, and, the allowed level of remaining storage space. If one of these levels exceed a certain threshold or the number of requests received for DI becomes equal or greater than the predefined threshold in a node, then the replication process starts, and function as follows: The PSM notifies the Replica Decider (RDec) components to start the replicating process. RDec creates query request RQ that contains the DI to be replicated, the reason for replication and some important indicators to help other nodes decide whether to accept or reject the RQ. After that, it sends RQ in the network. Each node receives this RQ, its PSM decides whether to place replicas by accepting RQ or to reject it according to its status and information included in RQ.

Discussion:

ZBR and RABC are group-based strategies. They replicate only the data which are requested, this can degrade the overall system performance because unpopular data are difficult to locate and will cause significant traffic overhead compared to the popular data. The both use the same method to maintain consistency. CreaM and data

replication are fully decentralized strategies. In data replication strategy, each node based only on access frequency and number of hops in order to take the decision of replication. In our opinion, it is an insufficient view that the issues that the replication must address (section II.B). Whereas, CreaM takes many parameter (CPU, energy, ect...) which makes CreaM, in our opinion, more efficient compared to the data replication strategy.

IV. REPLICATION STRATEGIES FOR MP2P

A. Context and Location-based Efficient Allocation of Replicas (CLEAR)[14]

CLEAR is an efficient group-based replication strategy to increase the data availability in MP2P networks. It exploits user mobility patterns and functions as follows: Periodically, SP receives a message from each mobile host (MH) which belongs to its group. This message contains four information related to each mobile node: the read-write log D-MH for each data item which MH stores, the read logs R-MH for each replica which MH stores, its available memory and its load status. All these information will be combined to create read-write D-SP logs and read logs R-SP for SP. The candidate data items from D-SP which should be replicated, that are whose their access frequency exceeds a certain threshold. SP selects the nodes where the data can be replicated depending on their load status, their available memory and the number of time they accessed the replicated data. It avoids selecting overloaded nodes or nodes with low memory. CLEAR takes into account other information which are MH's mobility patterns and the data item required by MH. This information is stored in the schedule and each MH sends its schedule to its SP as soon as it first enters. This information allows CLEAR to have better response times and increases data availability, by replication data item at MHs would soon come into the region and avoid replication at MHs that would quit the region.

B. A Collaborative replica allocation and deallocation approach (CADRE) [15]

CADRE tries to avoid thrashing conditions, where the mobile nodes expend more bandwidth and energy on allocating and deallocating the same data item than on answering queries. In order to achieve this, it performs allocation and deallocation decisions jointly among nodes in the network.

The other important goal of CADRE is to ensure fairness in replication by avoiding to give replication decisions only based on read/write ratio without considering the origin of queries. Thus, if a data item is accessed very frequently in the network, then it is replicated with high priority by creating several replicas, even if it is accessed (read) only by a single node. To do this, CADRE proposes a new parameter which is a score which is assigned for each data item to quantify its importance in the network as a whole. Thus, the data items which have a higher score are the first selected to be replicated, since these data items are important to the network as a whole.

CADRE is group-based strategy, the role of SPs is to execute replica allocation and deallocation collaboratively within the region that it covers and facilitates search in the network. Periodically, each mobile node sends its load status, energy status, available memory space status, the scores of items and the Flip Flop Ratio (FFR) values of the replicas stored at itself to the corresponding GN in its region. The

FFR is used for prevention of thrashing. To calculate it, each MH keeps track of the number of deallocations of each replica at itself over a period of time. If the FFR value of replica increases, then the probability of thrashing of this replica also increases. Hence, if the FFR for a replica exceeds a certain threshold, then it should not be deallocated. All these information allow the SP achieving fair replica allocation among the mobile nodes in the network with collaboration to the deallocation decisions.

C. Adaptive replication of images (Ada-Rep) [16]

Ada-Rep is group-based strategy. It can reduce the memory usage and request turnaround time. To decide the replica allocation of images, the SPs consider various parameters such as resources available, network state, trust values and average response time. Ada-Rep takes the decision for creating the new replica in parallel with lookup protocol and it operates as follows: Periodically, each peer informs its SP about all its queries executed successfully in that period. Then, The SP calculates the new average response time for each image transaction into the list of requested objects. The list of requested objects is maintained by each SP and contains the request count and average response of each requested object image. The new average response time for an object is compared to the previous one, if it is greater than, the SP authorizes the replication and allows to create more replicas for this object to increase its availability.

When the SP receives the requests from the requesting peer for object image, and if the desired replica count for that object does not attain. Then, it replies to the requesting peer with authentication certificate, the *replication permission* with the trust value and other information. With authentication certificate each requester peer can know information regarding level of access permission granted to it by its SP. After, the requesting peer sends the request with all this information to all the neighboring peers or the group of peers according to the types of searching methods used (Flooding or Multicast).

During the forwarding of request, the number of existing replicas of that image available in the cache memory of the nodes in that path is piggybacked in the message header of the request. When the source peer receives the request of image with *replication permission* and if the number of existing replicas of the image is less than the required number of replicas, then it launches the replication factor (RF) algorithm. The RF helps to locate the peer(s) to replicate the image. RF was calculated for each intermediate node except at the nodes already having the copies of that image on returning back and is stored in the message header with id's peers in the request back. RF is based on various parameters including memory, trust, bandwidth available and the processing power. When the requesting peer receives the request back, it chooses the intermediate peer(s) with the highest replication factors to store the replica of image. The replica can be the original image or its residue according to parameters. After, The requesting peer issues the final request with the replication array information to the source peer. The image/residue is replicated at the designated peers given in the replication array on the returning path.

D. The Random Replication with State Prediction Algorithm (RRSP)[17]

RRSP proposes a modified form of random replication. It uses the Self-Organizing Map (SOM) to predict the node's mobility state and to avoid replicating on the nodes that are about to leave from the network. The Self-Organizing Map (SOM) is the unsupervised learning, neural networks, it is used to classify input data into groups. The data are trained using the unsupervised learning where number inputs and outputs are specified.

According to random replication, when the request for an object is successful, the nodes which will store a copy of this object are selected randomly along the search path. Whereas RRSP uses the node's condition as the basis for choosing to replicate objects into these nodes. These nodes' condition is calculated by using the SOM algorithm. There are a 4-dimensional input vectors for SOM, Mean Time to Failure (MTTF), Current Uptime, Distance and Round Trip Time (RTT). To compute MTTF, each node recalls both of the last leave time and the last join time. Current uptime is the total time the node currently online. The distance between two nodes is obtained by getting the radio link length between them. The RTT is the return time measured by sending a packet from the local node to the remote node. There are two categories as output of SOM: 0 or 1. The number 1 for those who have enough time to stay in the network and 0 is for nodes about to leave. If the search for data succeeds, along the search path the SOM algorithm is applied for each node. Then, only the nodes whose their output set equal to 1 receive the copy of the requested data. RRSP minimizes number of replicas and therefore the storage space is diminished compared to random replication.

E. Randomized Distributed Algorithm (RDA)[18]

RDA aims to higher scalability and efficiency in a resource-limited network such as MP2P. Its main idea is to ensure that all query sets become k -coverable. This algorithm is executed by all nodes and it works as follows: Each node i records the hop-distance of objects in the distance vector $cost_i$ with size m , where m is the number of objects O in the network. During the period of t , each node i local-broadcasts the distance vector to its 1-hop neighbors. Whenever it receives all the distance vector of its 1-hop neighbors j , for each object O , if $cost_j(O) + 1 \leq cost_i(O)$, then $cost_i(O)$ will be updated to $cost_i(O) + 1$. After executing such operations for k times, if the value of each $cost_i(O)$ is smaller or equal to k , it indicates that the query set of all objects is k -coverable and the node i do nothing. Otherwise, node i chooses the candidate object for replicating/dropping with probability $(1 - P)$. If the local memory is full, then node i chooses the candidate object for dropping with $cost(O) = 0$, otherwise it will issue a request for replicating object O with $cost(O) > k$.

F. An Economic Model for Efficient Dynamic Replication (EcoRep)[19]

The main objectives of EcoRep are to ensure fair replica allocation, discourages free-riding and provides an incentive for users to become service-providers. For this purpose, it based on the data item's price to perform replica allocation. This price is paid by the requester node to the node that serves the request. On account of this value, EcoRep can be regarded as an incentive scheme because in order that MH will be able to issue its own request, it must provide service to the network. Additionally, it minimizes the

number of the free-riding nodes that do not participate in replication of data items.

Price of a data item relies on the number of users who accessed it, its access frequency, the number of its existing replicas, the average response time required for accessing it and its consistency. EcoRepis group-based strategy and does not require absolute replica consistency. Periodically, each peer calculates the price of its each item (original and replica), which were accessed by mobile nodes MNs from within the region of its corresponding super peer SP and by mobile nodes that are outside regions of its SP (external). All these prices must be sent to its corresponding SP with a list of data items (original or replica) stored at itself. When the SP receives price information from all the mobile nodes in its region, then it computes the total price of each item (original or replica). For the internally accessed items, SP sorts these items in descending. SP considers those items, whose prices exceed the average price, as candidates for replication. For the externally accessed items, the SP computes the price of each data item relative to every external SP and chooses the candidate item that should be replicated at MNs in the regions of other SPs. For each candidate item, SP sends a message to the relevant external SP, which will perform the replica allocation for this item with MNs within its region. For choosing the MNs that store the replicas, SP is based on more information sent by MNs within its region, which are: load, energy and available memory space status, current (x,y) coordinates of each MNs that will be used to estimate the network topology during the time of replica allocation.

G. Applicability of sub graph centrality to improve data accessibility among peers in MANETs (ASGC) [20]

The main aim of Pushpalatha et al. is to propose a data replication scheme that distributes the replica in a minimum number of nodes and let each mobile node to access the data in a minimum number of hops (one or two hops). This can decrease the data access delay, improve the response time and reduce the communication cost to access the data. In order to realize this, this technique uses mathematical concept known as minimum dominating set and sub graph centrality principle and consists of two phases: Initialization phase and maintenance phase. The node that holds the shared data is termed as data server and the nodes that access the data are termed as data clients.

In Initialization phase, the data servers that wish to replicate data to invoke the algorithm to find the minimum dominating set for the current topology [21]. The goal to construct a minimum dominating set will help the nodes to access the data in at most one or two hops. After identifying the dominating nodes (servers), the replica is distributed to those nodes. Dominating nodes, then broadcast the identifier of the data it holds to its neighbors in one or two hops. If the neighbor receives the broadcast message from one or more data servers, it stores the identifier of the data servers and the data identifiers in a table. This phase is suitable for dynamic network where the dominating nodes (servers) may move out of range and the client can not be access the data from data server in at most one or two hops. For this the maintenance phase uses sub graph centrality principle to identify a stable node and relocate the replica (choose new dominating nodes) if a server node moves out of range from its client.

Discussion:

CLEAR, EcoRep and CADRE are group-based strategies ; where the nodes communicate periodically their information (load status, memory, etc...) to the SP that it uses them to make the local decisions of replication (choosing the data to be replicated, the nodes where data can be replicated, etc...). The centralized nature of these solutions allows synchronized replication decisions between the nodes in the group, thus it avoids replica duplications and non-useful replica creation. Additionally, this allows making more accurate decisions (as decisions are based on a complete view of the vicinity). However, this can generate high communication cost, which is not suitable in such environment with very limited resources. CADRE and EcoRep is based on the CLEAR. EcoRep is the first work that proposed an economical model to avoid the problem of free riding. Ada-Rep is also a group-based strategy and SP are responsible for collecting the information from all other nodes in this group. The difference between Ada-Rep and CLEAR is that the SP is not responsible to choose the nodes where data can be replicated. These nodes are choosing in the request path based on some parameters (memory, trust, bandwidth available, etc...). Likewise, the data which to be replicated is only a digital image.

RDA is fully decentralized strategy in which the decisions made by each node only depend on its current state (the value of distance vector cost). It is completely different compared to strategies presented in this paper because it can be treated as Markov chain. However, this solution does not take in consideration the characteristics of Manet (mobility, the limited physical resources like bandwidth and energy, etc...). RRSP is also fully decentralized strategy. The main feature of this solution is the use of Self-Organizing Map to predict the node's mobility state. ASGC is the only technique which guarantees that each node can access the data in at most one or two hops and also it maintains minimum replica degree for each data item.

V. SUMMARY

MANETs are support environments for autonomous and decentralized applications like P2P file sharing. However, the dynamic and unpredictable nature of MANETs creates several challenges, which in turn impact the applications deployed on top of them. Increasing the data availability is the major challenge. To deal with this challenge, a number of data replication protocols have been proposed in the recent years. In this article, we explored some techniques that have a significant scientific contribution. In the following we try to draw some useful recommendations to take into consideration when we propose the replication strategies for mobile environment:

Wireless connections have limited bandwidth which must be reserved in priority for the applications. Consequently, the resource overhead imposed by running the replication system must be minimized.

The consistency maintenance is an important issue, and it must be addressed together with the replication technique in order to decrease the overhead of consistency maintenance and ensure that a data requester receive up-to-date data.

The main parameter for replication decision (to choose suitable data to replicate) is the access frequencies of data items.

The storage space is an important parameter to choose the suitable node to host replica. However, it is insufficient given that the node has other limitations like limited CPU, battery, bandwidth. Therefore, replication strategy must take into account these limitations.

Most of them assume that all mobile nodes collaborate fully in terms of sharing their memory space. In reality, however, some nodes may decide only to cooperate partially, or not at all, with other nodes. These nodes could then reduce the overall data accessibility in the network. Therefore, the replication must be applied, accompanied with an incentive mechanism.

Table I summarizes the comparison of replication strategies in mobile environment presented above in function of some criteria:

Type: strategy is group-based or is fully decentralized.

When: replication is executed in parallel with lookup protocol, periodically or when it is necessary (need). In the last case, the replication can be executed whenever new data is available in the system or when some parameter extends threshold like Access frequency.

Consistence issue.

Parameters considered: we cite all parameters considered in replication.

Advantages.

VI. CONCLUSION AND FUTURE WORK

Replication techniques are widely employed to improve the availability of data, enhancing performance of query latency and load balancing. In this paper, a state of the art of the existing replication techniques for Manet and M-P2P networks is presented. In our future work, we try to develop simultaneously data replication and data consistency maintenance methods and take into consideration recommendations that we presented in order to achieve high efficiency at a significantly lower cost

REFERENCES

- [1] "Gnutella," [retrieved: May, 2014]. [Online]. Available: <http://www.gnutella.com>
- [2] "Kazaa," [retrieved: May, 2014]. [Online]. Available: <http://www.kazaa.com>
- [3] "Skype," [retrieved: May, 2014]. [Online]. Available: <http://www.skype.com/>
- [4] "Seti@home," [retrieved: May, 2014]. [Online]. Available: <http://setiathome.berkeley.edu/>
- [5] E. Cohen and S. Shenker, "Replication strategies in unstructured peer-to-peer networks," 2002, pp. 177–190.
- [6] S. M. Thampi and K. C. Sekaran, "Review of replication schemes for unstructured p2p networks," CoRR, vol. abs/0903.1734, 2009.
- [7] R. moufida and B. Mahfoud, "A Comparative Study of Replication Schemes for Structured P2P Networks", in ICIW 2014, The Ninth International Conference on Internet and Web Applications and Services, 20 July 2014, pp. 147-158.
- [8] T. Hara, "Effective replica allocation in ad hoc networks for

improving data accessibility,” pp. 1568–1576, 2001. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=916653>

[9] K. Rajeev and P. Kumar, “Replica allocation technique based on clusters for manets,” in International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE’2012), March 24-25 2012.

[10] K. Shi, R. Chen, and H. Jin, “Zone-based replication scheme for mobile ad hoc networks using cross-layer design,” in MSN’06, 2006, pp. 698–710.

[11] T. Clausen and P. Jacquet, “Optimized Link State Routing Protocol (OLSR),” RFC 3626 (Experimental), Internet Engineering Task Force, October 2003. [Online]. Available: <http://www.ietf.org/rfc/rfc3626.txt>

[12] M. Samira, G. Mohamed, and B. Nadjib, “Data replication in mobile ad hoc networks,” in Mobile Ad-hoc and Sensor Networks, vol. 4325, 2006, pp. 685–697.

[13] Z. Torbey, N. Bennani, D. Coquil, and L. Brunie, “Cream: User-centric replication model for mobile environments,” in International Workshop on “Mobile P2P Data Management, Security and Trust (M-PDMST2010)” in conjunction with the 11th International Conference on Mobile Data Management (MDM 2010), IEEE, Ed., May 2010, pp. 348–353. [Online]. Available: <http://liris.cnrs.fr/publis/?id=4614>

[14] A. Mondal, S. Kumar, and M. M. Kitsuregawa, “Clear: An efficient context and location-based dynamic replication scheme for mobile-p2p networks,” in Proc. DEXA, 2006, pp. 399–408.

[15] A. Mondal, S. K. Madria, and M. Kitsuregawa, “Cadre: A collaborative replica allocation and deallocation approach for mobile-p2p networks.” in IDEAS. IEEE Computer Society, 2006, pp. 21–28. [Online]. Available: <http://dblp.uni-trier.de/db/conf/ideas/ideas2006.htmlMondalMK06>

[16] A. R. Kumar, K. M. Sanjay, and H. Takahiro, “Adaptive searching and replication of images in mobile hierarchical peer-to-peer networks,” Science Direct, 30 January 2007.

[17] G. C. Soriano and Y. Urano, “Replication with state using the self-organizing map neural network,” pp. 383–388, 2011.

[18] H.-Z. Chou, S.-C. Wang, and S.-Y. Kuo, “A randomized distributed algorithm for peer-to-peer data replication in wireless ad hoc networks.” in PRDC. IEEE Computers Society, 2007, pp. 163–170. [Online]. Available: <http://dblp.uni-trier.de/db/conf/prdc/prdc2007.htmlChouWK07>

[19] A. Mondal, S. Madria, and M. Kitsuregawa, “Ecorep: An economic model for efficient dynamic replication in mobile-p2p networks.” in COMAD, L. V. S. Lakshmanan, P. Roy, and A. K. H. Tung, Eds. Tata McGraw-Hill Publishing Company Limited, 2006, pp. 116–127. [Online]. Available: <http://dblp.uni-trier.de/db/conf/comad/comad2006.htmlMondalMK06>

[20] [M. Pushpalatha](#), [T. Ramarao](#) and [Revathi Venkataraman](#), “Applicability of sub graph centrality to improve data accessibility among peers in MANETs”, in Peer-to-Peer Networking and Applications, vol. 7, June 2014, pp. 129-146.

[21] Narsingh Deo “Graph Theory with Applications to Engineering and Computer Science”, Prentice Hall, 1974.

TABLE I. COMPARISON BETWEEN DIFFERENT REPLICATION STRATEGIES IN MOBILE ENVIRONMENT

Replication techniques	Criteria	Type	When	Consistence issue	Parameters considered	Advantages
MANET	RABC	Group-based strategy	Parallel with lookup protocol	Yes	Storage space	Is simple to deploy
	ZPR	Group-based strategy	Parallel with lookup protocol	Yes	Storage space	Improves the overall performance of data access
	Data Replication	Fully decentralized strategy	It is necessary	Non	Access frequency	Improves the overall performance of data access
	CReaM	Fully decentralized strategy	It is necessary	Non	Level of CPU, battery and storage space + Access frequency	Increases the data availability in a significant way, with high rate of user satisfaction
MP2P	CLEAR	Group-based strategy	Periodically	Yes	Read-write log, storage space, mobility patterns and load status	Exploits mobility pattern which allows to facilitating better resource utilization, likely better query response times
	CADRE	Group-based strategy	Periodically	Non	Load status, energy status, storage space and FFR value	Avoids thrashing conditions by performing allocation and deallocation decisions jointly
	Ada-Rep	A mix between Group-based and fully decentralized strategy	It is necessary	Non	Bandwidth, storage space, processing power, trust value and average response time	is the best strategy to search and replicate images
	RRSP	Fully decentralized strategy	It is necessary	Non	MTTF, current uptime, distance and RTT	Increases performance MP2P compared to random replication by using the SOM technique
	RDA	Fully decentralized strategy	It is necessary	Non	Storage space and cost value	Is completely different compared to other strategy presented in this paper. It is operated as a Markov chain
	EcoRep	Group-based strategy	Periodically	Yes	Price: number of users who accessed, access frequency, number of existing replicas and , the average response time + load, energy and available storage space status and current coordinates	Ensures fair replica allocation and discourages free-riding
	ASGC	Fully decentralized strategy	It is necessary	Non	Mobility	Each node can access the data in at most one or two hops and also it maintains minimum replica degree for each data item.

Comparaisons des approches d'interopérabilité des réseaux P2P

Hadj Henni M'hamed

Department of Computer Science, Hassiba Ben Bouali
University, Chlef, Algeria

m.hadjhenni@univ-chlef.dz, hadjhenni@gmail.com

BENCHAÏBA Mahfoud

LSI, Dep. of Computer Science, USTHB, B.P. 32 El-Alia,
Bab Ezzouar- Algiers – Algeria,
benchaiba@lsi-usthb.dz

Résumé

Les réseaux p2p connaissent un grand succès et une utilisation élargie. Néanmoins, vu que ces réseaux p2p utilisent des interfaces propriétaires et fournissent des services différents ; cela fait que ces réseaux sont isolés les uns par rapport aux autres. Pour résoudre cette problématique, des approches ont été proposées pour rendre des réseaux p2p différents interopérables. Malgré cela, ces approches restent non utilisées et non évaluées.

A notre connaissance, il n'existe pas des travaux qui ont étudiés ces différentes approches d'interopérabilités des réseaux p2p, ou qui ont essayé de comparer entre ces différentes approches, cet article vient pour répondre à ces besoins.

Cet article, on s'est proposé de définir l'interopérabilité en réseaux p2p, exposé et analysé ces approches pour déterminer leurs points forts et leurs points faibles, comparé ces approches. Cette étude nous a permis de cerner l'état d'avancement des recherches sur l'interopérabilité des réseaux p2p, et d'évaluer ces approches.

Mots clés—Réseaux Peer to peer, Interopérabilité.

I. INTRODUCTION:

Les réseaux p2p ont connu un grand succès durant cette dernière décennie. Cela s'est traduit par le développement de plusieurs applications p2p offrant différents services. Ces applications sont souvent propriétaires. L'utilisateur désirent bénéficier des différents services existant, doit installer toutes ces applications et rejoindre leurs réseaux p2p correspondants, ce qui est fastidieux voire impossible.

D'ici on ressent le besoin des utilisateurs d'une application donnée de pouvoir accéder et bénéficier des services d'autres applications p2p sans avoir à les installer. Pour se faire on doit rendre l'application de l'utilisateur interopérable avec les autres applications p2p. Dans ce sens, les chercheurs ont proposé plusieurs approches pour rendre les réseaux p2p interopérables. Malgré la différence de leurs dates d'apparition, les nouvelles approches ne font pas référence les unes aux autres. De plus, il n'y a pas eu des travaux de recherche qui visent à évaluer et comparer ces approches. Dans cet objectif, on essaiera dans cet article d'analyser et de comparer ces différentes approches d'interopérabilité des réseaux p2p.

Cet article est organisé comme suit. La section 2 présente un préliminaire sur la notion d'interopérabilité en général puis l'interopérabilité en réseau p2p est présenté en section 3. Les approches d'interopérabilité ainsi que leurs analyses feront l'objet de la section 4. On effectuera la comparaison dans la section 5 et on termine par une synthèse, et une conclusion en section 6 et 7 respectivement.

II. PRELIMINAIRES

A. Interopérabilité des systèmes

Il y a une nécessité d'une définition précise de l'interopérabilité, car le terme peut avoir différentes interprétations dans des contextes différents [1].

De nombreuses définitions existent dans la littérature (pour n'en citer que les plus communes) [1] :

L'interopérabilité est la capacité que possèdent deux ou plusieurs systèmes ou composants à échanger des informations puis à exploiter les informations venant d'être échangées [2].

L'interopérabilité est la capacité à communiquer, exécuter des programmes, transférer des données entre différentes unités fonctionnelles de manière à ce qu'un utilisateur n'ait besoin que de peu ou pas de connaissances sur les caractéristiques de ces unités [3].

La capacité des systèmes, des unités, ou des forces d'accepter et de fournir des services à d'autres systèmes, unités, ou forces, et d'employer les services ainsi échangés pour leur permettre de fonctionner efficacement ensemble [4].

III. INTEROPERABILITE EN RESEAUX P2P

Il n'existe pas de définition dans la littérature de l'interopérabilité dans les réseaux p2p, pour cela on a proposé la définition suivante :

Le terme interopérabilité signifie les mécanismes mis en œuvre pour assurer d'une manière transparente la communication et le partage des services et des ressources entre des applications p2p appartenant au même réseau ou à des réseaux p2p différents en terme d'architecture de réseaux p2p, méthodes de routage, messages de communication et services proposés (voir Figure 1).

A. Quand se pose le problème d'interopérabilité entre réseaux p2p ?

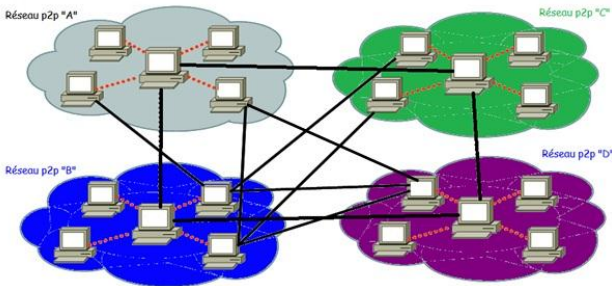
Le problème de l'interopérabilité entre applications p2p se pose fortement dans des réseaux p2p hétérogènes à cause de :

La diversité des architectures de réseaux p2p (centralisée, décentralisée, hybride, structuré).

La diversité des méthodes de recherches de ressources. (Recherche par index, par inondation ou par hachage de clé).

La diversité des types de messages de communication.

La diversité des types de services assurés par chaque application (partage de fichiers, I. M, streaming, ...).



Exemple de réseaux p2p différents en communication.

B. Intérêt de l'interopérabilité en réseau p2p

Les applications p2p proposent divers services et divers avantages, l'intérêt de rendre des applications p2p différentes interopérables est de permettre de :

Combiner et partager les fonctionnalités intéressantes dans chaque réseaux p2p (efficacité de méthodes de routage, de méthode de recherche,..)

Elargir le nombre de ressources disponibles et faciliter à l'utilisateur l'accès à toutes les ressources partagées à partir d'une seule application. Sans l'interopérabilité, l'utilisateur sera obligé d'utiliser plusieurs applications p2p et se connecté aux réseaux correspondants pour avoir toutes les réponses correspondantes à sa requête, ce qui est fastidieux et presque impossible.

Répliquer des services et des données sur de nouveaux réseaux p2p ; ce qui garantit la disponibilité des ressources (même en cas d'indisponibilité de l'original) et l'accélération de téléchargement des ressources.

Simplifier l'interconnexion et assurer la communication entre les peers de systèmes p2p différents.

IV. APPROCHES D'INTEROPERABILITES EN RESEAUX P2P

Les applications courantes de p2p sont généralement construites sur une plate-forme spécifique de p2p. Des applications p2p différentes offrant des services similaires ne peuvent pas partager directement des données et des services les unes avec les autres.

Des approches assurant l'interopérabilité entre des réseaux p2p différents ont vu le jour tel que: l'approche de Gateway, l'approche PeerBus, l'approche Proxy, l'approche d'index structuré distribué et l'approche d'interconnexion H-p2pSIP. Dans ce qui suit, nous allons exposer, analyser et comparer ces approches.

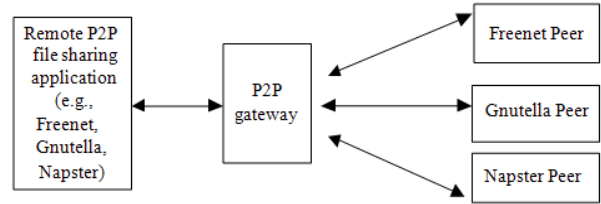
A. Approche de Gateway (passerelle)

Cette approche est établie autour de trois applications p2p parmi les plus populaires en partage de fichiers et qui sont : Freenet, Gnutella, et Napster. Ces applications viennent avec des caractéristiques différentes. Freenet et Gnutella utilisent l'architecture p2p, Napster utilise l'architecture p2p centralisée via un serveur central pour gérer la découverte par les pairs et la recherche de contenu [5].

Cette approche propose un framework qui permet l'intégration de protocoles de partage de fichiers en utilisant

une passerelle (Gateway) p2p, qui reçoit les messages des différents protocoles et les convertis dans le protocole spécifique de l'application de partage de fichiers distante.

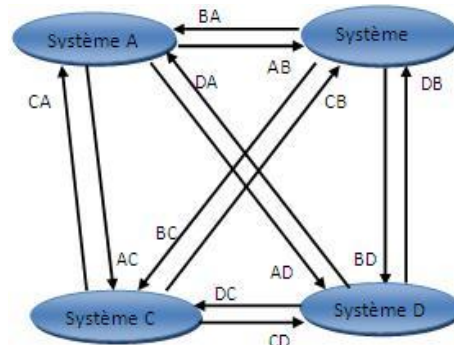
En outre, tous les messages sortants de chaque protocole seront convertis vers tous les autres protocoles et ensuite envoyés sur le réseau. L'architecture de la mise en œuvre de la passerelle p2p est donnée à la Figure 2.



Framework d'interopérabilité des protocoles p2p de partage de fichiers [5].

Analyse de l'approche Gateway :

Comme le montre la Figure 3: la conversion des messages entre quatre systèmes choisis est de 12 messages. En général pour rendre n réseaux p2p interopérables, on aura un graphe complet et on aura besoin d'échanger n*(n-1) messages.



La conversion des messages dans l'approche Gateway.

Avantages

L'interopérabilité est assurée.

Echange transparent de messages.

Inconvénients

Cette approche présente une expansion limitée et entraîne plusieurs limitations, telles que la nécessité d'une connectivité constante à tous les réseaux soutenus, le manque de métadonnées pour identifier des contenus partagés similaires, l'exigence d'une structure d'indexation centralisée [7].

Pour intégrer un nouveau système p2p au Gateway, on doit maîtriser ses messages et assurer leurs conversions vers tous les autres protocoles p2p déjà interopérables.

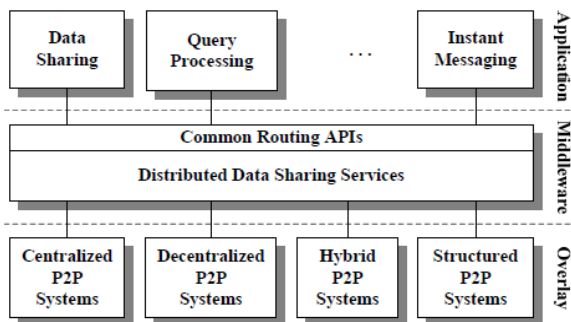
Surcharge du réseau, vu que le même message sera dupliqué et envoyé dans les différents autres protocoles.

B. Approche du middleware (intergiciel) PeerBus

Cette approche définit un middleware p2p qui fait abstraction de la complexité et l'hétérogénéité des systèmes sous-jacents p2p avec leurs multitudes de services, de protocoles et d'applications. PeerBus se concentre sur

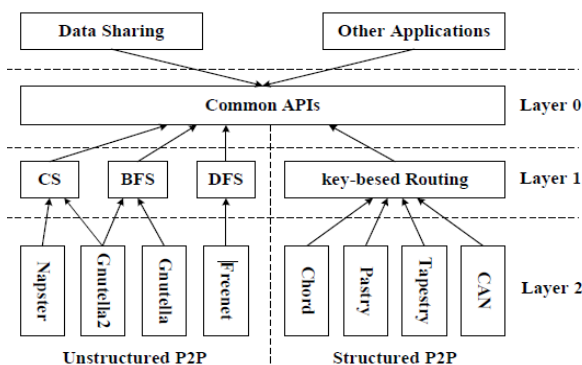
l'abstraction des opérations courantes de partage de données (en général partage de fichiers) [6].

La couche middleware de PeerBus est composée d'API communes et des services distribués de partage de données (voir Figure 4).



L'architecture de PeerBus [6].

La Figure 5 illustre une abstraction hiérarchique des services de partage de données p2p.

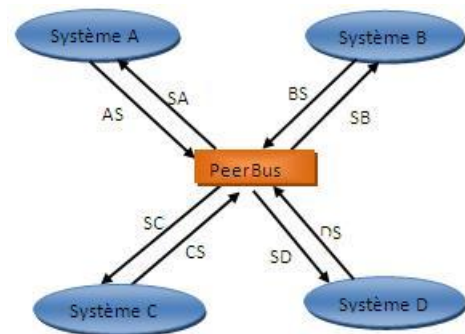


Abstraction des services de partage de données p2p [6].

Les applications p2p accèdent aux réseaux p2p sous-jacents comme un réseau virtuel à travers les API communes (couche 0). De cette manière, l'interopérabilité entre les différents systèmes p2p peut être réalisée en utilisant les services communs de l'API, qui est la partie centrale de la couche middleware de PeerBus. Les services de partage de données basés sur la recherche centralisée, ou bien sur la recherche en largeur d'abord (BFS : Breadth First Search) ou sur la DHT (Distributed Hash Table) fournissent les abstractions de haut niveau pour les couches sous-jacentes de systèmes p2p [6].

Analyse de l'approche PeerBus

Comme le montre la Figure 6: la conversion des messages entre quatre systèmes choisis est de 08 messages ; en général pour rendre n réseaux p2p interopérables à l'aide de PeerBus on a besoin d'échanger $2*n$ messages. Cela signifie que cette approche est plus efficace que l'approche Gateway.



La conversion des messages dans l'approche PeerBus.

Avantages

Assure l'interopérabilité.

L'expansion est possible (pour rendre un nouveau réseau p2p interopérable avec les réseaux déjà existants, il suffit de traduire ses messages vers les messages de Peerbus et inversement de traduire les messages de Peerbus vers les messages de ce nouveau réseau).

PeerBus permet l'interopérabilité entre des systèmes p2p différents, y compris les systèmes p2p non structurés et structurés [6].

En tant que middleware, PeerBus est indépendant des systèmes sous-jacents p2p par abstraction de services communs à partir de différents systèmes p2p [6].

Réduction du nombre de messages échangés pour assurer l'interopérabilité par rapport aux autres approches.

Inconvénients

Le téléchargement multi sources n'est pas pris en charge.

Le manque de métadonnées pour identifier des contenus partagés similaires.

C. Approche de Proxy (passerelles)

Cette approche montre comment, par des moyens de proxy et de correspondance sémantique, il est possible de connecter un réseau P2P comme Edutella avec d'autres systèmes extérieurs à ce réseau [7].

Edutella [8] est un réseau p2p dans lequel les objets d'apprentissage (L.O : Learning Object) sont offerts gratuitement et tout le monde est capable de se joindre à ce réseau. Edutella est construit sur la base de la plateforme JXTA.

Afin que plusieurs entités puissent communiquer, il est nécessaire qu'ils s'entendent sur un protocole commun et sur une interface commune. Cette approche est construite sur la base :

Des méthodes spécifiées dans l'initiative d'interface simple de requête (S.Q.I : Simple Query Interface), une norme qui mûrit rapidement, en utilisant sa liaison de service Web [7].

Du langage QEL (Query Exchange Language) [9]. Dans ce système, plusieurs emballages sont mis en œuvre en vue de fournir un accès aux entrepôts les plus courants (bases de données relationnelles ou de systèmes de fichiers RDF, etc.). Le système reçoit

une requête dans QEL et la convertit vers le langage de requête locale [7].

Interopérabilité avec proxy

Pour garantir l'interopérabilité, cette approche utilise les Proxy qui sont utilisés pour connecter des pairs dans un réseau p2p avec le monde «extérieur» [7].

Aujourd'hui, de nombreux systèmes fournissent leurs services et ressources via des services Web et les Proxy de cette approche sont conçus pour faire communiquer le protocole JXTA (qui est la base du réseau p2p Edutella) et les autres interfaces via un protocole de service Web basé sur l'interface de requête simple (SQI) [7].

Dans cette approche, on peut prévoir deux scénarios possibles[7]:

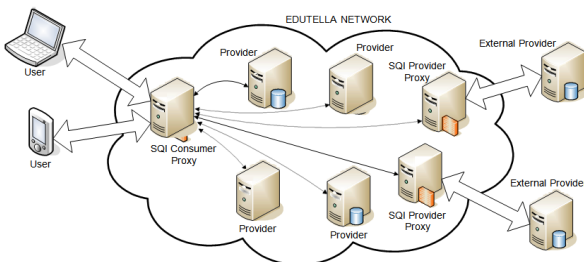
Un consommateur externe (client) au réseau p2p (le réseau p2p est sollicité) veut interroger le contenu dans le réseau p2p. Par exemple, supposons qu'on aimerait offrir le contenu d'un réseau p2p sur un site web. La solution est de transmettre la requête à partir du site Web vers le réseau p2p, par exemple, au moyen des Services Web et de récupérer la réponse avec le même mécanisme.

Un fournisseur externe veut offrir un contenu sur le réseau p2p(le réseau p2p via un peer sollicite un autre système). On suppose que les fournisseurs qui ont déjà mis en œuvre une interface basée sur les services Web n'auront pas envie de dépenser le temps et de l'argent dans le développement d'interface propriétaire du réseau. Au contraire, ils aimeraient réutiliser celle qu'ils ont, ce qui faciliterait également son administration.

Selon ces deux scénarios (voir Figure 7), il y a deux types différents de proxy. Le premier scénario fait appel à " un proxy consommateur " et le dernier à un "proxy fournisseur". Un proxy consommateur agit comme un médiateur entre un client externe qui veut interroger le réseau et le réseau p2p lui-même. Un proxy fournisseur agit comme un médiateur afin de fournir le contenu d'un fournisseur externe au profit d'un réseau p2p [7].

Mappages sémantiques (Semantic Mappings)

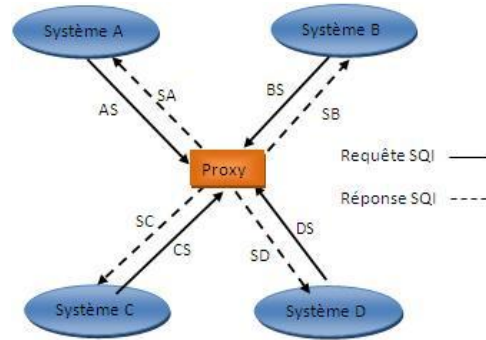
Afin d'assurer l'interopérabilité sémantique dans le réseau Edutella, les auteurs ont mis au point un module de réécriture de requêtes (Query Rewriting Module) qui transforme une requête q1 en une requête q2 en fonction de l'ensemble des mappages spécifiées. Ce module est conçu pour fonctionner sur des couples de correspondances sans un schéma unifié ou dans les approches d'intégration selon un schéma local (Local As View) [7].



Architecture de l'approche de proxy [7].

Analyse de l'approche Proxy

Comme le montre la Figure 10, pour rendre quatre systèmes p2p interopérables nous avons besoin d'échanger 08 messages. En général, pour rendre n réseaux p2p interopérables on a besoin d'échanger 2*n messages. Cela signifie que cette approche est plus efficace que l'approche Gateway.



La conversion des messages dans l'approche Proxy.

Avantages

L'Interopérabilité est assurée.

Ouverture de réseau p2p pour le reste du monde en utilisant les Proxies.

Utilisation de métadonnées pour identifier des Contents partagés similaires.

L'implémentation de cette approche existe.

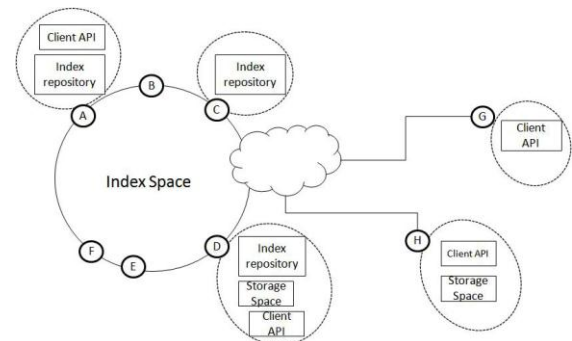
Inconvénients

Cette approche permet l'interopérabilité seulement avec les réseaux p2p à base de JXTA.

En cas de panne de l'un des proxies, l'interconnexion de réseau p2p et l'environnement extérieur sera infaisable.

D. Approche d'index structuré distribué pour le partage de données

Cette approche est axée sur la création d'un index global dans un espace de stockage entièrement distribué qui détient des métadonnées structurées sur les données à partager. En outre, cette approche fournit des mécanismes pour l'indexation et la recherche des données, rendant possible l'accès aux données provenant des différents systèmes et applications, offrant ainsi les avantages de partage peer-to-peer de fichiers à des systèmes quelconques (réseaux p2p, serveur web, serveur ftp,...) [10].



Architecture du système à base d'index [10].

Cette approche se base sur la table de hachage distribuée BambooDHT qui a été adaptée pour être utilisée comme la

base de l'index pleinement distribué. En plus, une API a été développée pour implémenter les fonctions de base d'accès à l'index ainsi que le soutien des extensions de nouveaux utilisateurs ou de nouveaux systèmes spécifiques, visant à fournir l'accès à n'importe quelle application [10].

Le modèle proposé est divisé en trois modules principaux: l'espace d'Index, l'espace de stockage et l'API client. Chaque nœud contient un ou plusieurs modules du système [10].

Comme le montre la Figure 9, le système proposé est très modulaire, car les nœuds participants peuvent prendre un ou plusieurs rôles dans le réseau créé [10].

L'espace d'index est un système de stockage qui distribue d'une manière transparente les métadonnées concernant les fichiers partagés sur un espace distribué, fondée sur la participation des nœuds dans un réseau p2p structuré. Les indices des fichiers partagés sont stockés dans ce module sous la forme de paires $[x, v]$, où x est l'une des balises utilisées pour rechercher un fichier, et v son descripteur sérialisé en format RDF. Le descripteur d'un fichier contient des informations comme le prénom et le nom de l'auteur, le titre du fichier, le format, la taille, l'année de publication et un ou plusieurs emplacements où il peut être trouvé [10].

Le module de l'espace de stockage est responsable du stockage des fichiers partagés et offerts par l'intermédiaire des protocoles tels que HTTP, FTP, BitTorrent, etc.... [10].

Le module API client est fourni sous forme d'une logithèque qui contient les fonctions de base pour accéder au dépôt. Cette API fournit une couche d'abstraction des mécanismes de téléchargement des fichiers, ce qui permet aux utilisateurs de sélectionner une méthode de téléchargement personnalisée en cas de besoin [10].

Analyse de l'approche d'index structuré distribué

Cette approche n'utilise pas de conversion de messages, car tous les messages échangés sont des messages H-p2pSIP. Les super peers jouent le rôle d'aiguilleurs des messages vers un autre peer du même domaine ou vers un autre super Peer du niveau interconnexion, si la ressource appartient à un domaine différent.

Avantages

Ouverture de réseau p2p au reste du monde on utilisant l'index distribué.

Evite de dépenser le temps et l'argent dans le développement d'interface propriétaire du réseau.

Visibilité de toutes les sources qui détiennent une copie de la ressource recherchée.

Gain de temps : Au lieu de chercher les ressources, on consulte directement l'index.

L'index réalisé bénéficie des avantages des réseaux p2p structurés (évolutivité, robustesse, fiabilité et organisation).

Les résultats de recherches sont plus fructueux vu qu'on peut utiliser plusieurs descripteurs dans la recherche (nom d'auteur, mot clé, ...etc.) au lieu de se limiter seulement au nom du fichier.

Cette approche améliore la disponibilité des données, en offrant l'interopérabilité entre les différents

protocoles de communication de réseaux et entre les réseaux p2p [10].

L'expansion est possible (un système quelconque peut référencier ces ressources sur l'index distribué).

Utilisation de métadonnées pour identifier des Contenus partagés similaires.

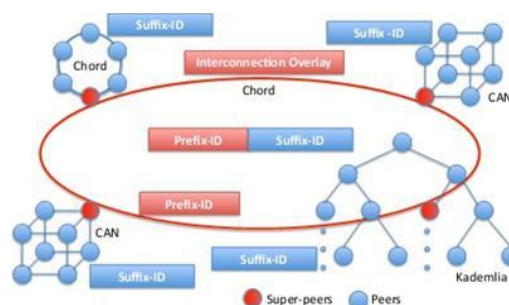
Inconvénients

L'index fait référence à des ressources de plusieurs types ; alors que l'utilisateur peut se restreindre à un type donné (media, logiciel, service,...) ce qui augmente le temps de recherche d'une ressource.

E. Approche d'interconnexion de réseaux peer-to-peer structurés avec une topologie hiérarchique

Cette approche permet de rendre interopérable des réseaux p2p structurés et cela à l'aide de H-p2pSIP qui définit un réseau logique composée de deux niveaux de hiérarchie, un exemple est donné dans la Figure 12. Le but de cette hiérarchie à deux niveaux est de permettre l'échange d'informations entre les différents domaines de réseaux DHT. Le niveau inférieur est composé de différents domaines qui veulent échanger des informations entre eux. Chaque domaine est indépendant des autres et, par conséquent, les pairs dans chaque domaine implémentent un réseau DHT selon les préférences du domaine. Ainsi, le réseau DHT peut être différent d'un domaine à autre. D'autre part, le niveau supérieur est composé d'un seul réseau logique DHT, nommé réseau logique d'interconnexion qui offre un service semblable à un service d'annuaire entre les différents domaines depuis le niveau inférieur de la hiérarchie. Le but de ce niveau d'interconnexion est d'acheminer les requêtes entre les différents domaines lorsqu'un pair d'un domaine veut récupérer des informations stockées dans un autre domaine. Ce niveau d'interconnexion peut être basé sur n'importe quel réseau logique DHT [11].

Afin de soutenir ces fonctionnalités, H-p2pSIP utilise un espace hiérarchique des identifiants composés de Hierarchical-IDs. Un identifiant hiérarchique-ID contient deux identifiants concaténés: un préfixe-ID et un suffixe-ID [11].



Architecture du système à base d'index [11].

Le préfixe-ID est utilisé pour acheminer (router) les requêtes dans le niveau d'interconnexion entre les différents domaines. Cela implique que tous les pairs et toutes les ressources appartenant au même domaine partagent le même préfixe-ID. D'autre part, le suffixe-ID est utilisé seulement dans le domaine auquel appartient un pair (peer) et permet de localiser n'importe quelle ressource dans le réseau de ce domaine. Ainsi, cette conception permet le routage des requêtes entre les différents domaines. Lors de la recherche d'une ressource dans un autre domaine, la requête est dirigée

vers le domaine souhaité en utilisant le préfixe-ID. Enfin, la ressource désirée dans le domaine externe est trouvée par l'intermédiaire du suffixe-ID [11].

Analyse de l'approche H-p2pSIP

Cette approche n'utilise pas de conversion de messages, car la recherche se fait par accès à l'index.

Avantages

Assure l'interopérabilité entre des réseaux p2p structurés.

Les applications p2p peuvent accéder aux ressources des différents systèmes p2p de manière transparente.

L'implémentation de cette approche existe.

Inconvénients

H-p2pSIP permet l'interopérabilité seulement entre des systèmes p2p structurés.

Ne prend pas en charge le téléchargement multi sources.

Le manque de métadonnées pour identifier des Contenus partagés similaires.

F. L'approche à base de la plateforme JXTA

Divers projets P2P sont disponibles sur Internet, ils offrent des services de recherches distribuées temps réel (ex : Distributed Hash Table, Query Flooding, ...), des services de mise de partage temporaire d'information (ex : Gnutella[12], eDonkey[13], Kazaa[14], ...), des services de calculs distribués (ex : SETI@home[15], ...), des services de messagerie instantanée (ex : Jabber[16], MSN Messenger[17], ICQ[18], ...), ou des services de distribution du stockage de fichiers (CFS [19], PAST [20], OceanStore [21], Ivy [22], Freenet [23], ...). Ces services sont conçus de manière à, utiliser au mieux les ressources disponibles sur l'Internet telles que la bande passante, le temps de calcul et l'espace de stockage.

Malgré leurs bonnes intentions, ces solutions pèchent par leur hétérogénéité et leur non-interopérabilité. Pour cette raison, Sun Microsystems a proposé un framework P2P pour tenter de résoudre ce problème d'hétérogénéité et de non-interopérabilité lié aux diverses implémentations P2P : le framework JXTA. Au commencement, ce framework était fermé mais, par la suite, Sun a décidé de le rendre Open Source de manière à, le rendre plus populaire auprès des développeurs et utilisateurs en vue d'en assurer le succès [24].

Le but du Framework JXTA est de créer une plateforme qui permet de construire un grand ensemble de services et d'application distribuées pour tout dispositif qui pourrait être un peer. JXTA permet, entre autre, aux développeurs de se concentrer sur le développement de leurs applications [24].

L'interopérabilité constitue le premier objectif du framework JXTA et vise essentiellement à simplifier l'interconnexion, la localisation et la communication entre les peers d'un système P2P. L'idée sous-jacente est de permettre à ces peers d'offrir de manière la plus transparente possible des services et de participer plus aisément aux activités proposées par les différentes communautés P2P [24].

Analyse de l'approche JXTA

Le but de JXTA est donc de devenir un standard dans ce domaine évitant de plus, de réinventer la roue à chaque

développement. Et par là en rend tous les applications à base de JXTA interopérable. JXTA n'est pas une approche d'interopérabilité, mais parmi ses objectifs et ses points forts on trouve l'interopérabilité.

Cette approche n'utilise pas de conversion de messages, car toutes les applications sont conçues à base de JXTA.

Avantages

Assure l'interopérabilité entre les applications p2p basé sur JXTA seulement.

prise en charge de plusieurs services de réseaux p2p (partage de fichiers, Messagerie Instantanée,....)

Inconvénients

l'interopérabilité se confine aux applications p2p basé sur JXTA.

Ne prend pas en charge les réseaux structurés.

Le manque de métadonnées pour identifier des Contenus partagés similaires.

V. COMPARAISON ENTRE LES APPROCHES D'INTEROPERABILITES DES RESEAUX P2P

D'après nos études sur l'interopérabilité des réseaux p2p, nous avons constaté que ce domaine est peu étudié. De plus, il n'existe pas d'études qui ont évaluées ces différentes approches d'interopérabilité de réseau p2p. Pour ces raisons, on a ressenti le besoin de comparer les différentes approches d'interopérabilité de réseau p2p.

A. Critères de comparaison

La majorité des approches d'interopérabilité de réseaux p2p n'ont pas d'implémentations ce qui rend la tâche de leur comparaison très difficile. Pour pouvoir établir des comparaisons entre ces différentes approches, nous proposons de prendre en compte les critères suivants qui sont en majorité en relation avec les caractéristiques des réseaux p2p:

Type d'architecture : Architecture p2p décentralisée, architecture p2p centralisée, architecture p2p hybride (avec Super-peers), et architecture structurée (utilisation de DHT).

Types de services des réseaux p2p à rendre interopérables (partage de fichiers, Messagerie Instantanée, Streaming,...).

Méthodes de recherche utilisées dans les réseaux p2p tel que : Les index centralisés (architecture p2p centralisée) ; Les index distribuée (architecture p2p hybride) ; modèle d'inondation de requêtes (architecture p2p décentralisée) ; modèle de routage de documents (architecture structurée).

Nombres de messages échangés entre n systèmes p2p pour assurer l'interopérabilité entre eux.

Type de téléchargement de fichiers: Téléchargement à partir d'une source unique, téléchargement multi-sources et téléchargement segmenté multi-sources.

Interopérabilité avec d'autres systèmes non p2p (par exemple avec serveur web ou avec serveur ftp).

Existence d'implémentation : Existence d'implémentation et d'évaluation de l'approche proposée.

Comparaison

La comparaison (Tableau I) a révélé que certaines approches prennent en charge tous les types d'architectures, certaines d'autres sont conçues pour des architectures spécifiques telles que l'approche proxy qui est prévu seulement pour les réseaux à base de JXTA et l'approche H-p2pSIP qui prend en charge seulement les réseaux p2p structurés.

Le service pris en charge par ces différentes approches d'interopérabilité est le service de partage de fichier ; seule l'approche H-p2pSIP en plus du partage de fichier, elle prend en charge le service de messagerie instantanée (M.I).

Concernant le nombre de messages échangés pour garantir l'interopérabilité entre N systèmes p2p, et d'après le Tableau I ; on constate que l'approches PeerBus et l'approche Proxy sont très intéressantes vis-à-vis du nombre de messages nécessaires pour garantir l'interopérabilité, car ces approches font des conversions vers un standard qui joue le rôle de pivot.

Le type de téléchargement utilisé dans la plupart des approches est le téléchargement d'une source unique, alors que le téléchargement segmenté multi-sources est assuré seulement par l'approche Index, ce type de téléchargement est très intéressant pour améliorer le temps de téléchargement.

Certaines approches sont conçu pour assurer l'interopérabilité entre seulement des réseaux p2p, alors que d'autres approches assure en plus de cela l'interopérabilité entre des systèmes p2p et d'autres systèmes (Web, ftp,...).

La majorité des approches d'interopérabilités de réseaux p2p n'ont pas été implémenté, à l'exception de l'approche Proxy et l'approche H-p2pSIP, ce qui rend l'évaluation des autres approches non implémenté difficile.

VI. SYNTHÈSE

Cette étude nous a permis d'explorer l'interopérabilité de réseaux p2p ; ainsi que les différentes approches proposées ; nous avons analysés ces différentes approches ; d'après le tableau qui résume la comparaison entre ces différentes approches et on favorisant les critères d'existence d'implémentation et de types d'architecture on peut dire que les approches les plus intéressantes sont : l'approche des

index ; vu qu'elle permet de savoir si une ressource existe dans tous les réseaux à l'aide de l'index distribué et cela avant de solliciter la ressource ; après cela vient l'approche HP2PSIP ; qui définit deux niveaux d'interconnexion ; le premier inter domaine et le deuxième au niveau de chaque domaine ; cette approche prend en charge les réseaux mobiles ; mais ne prend en charge que l'interopérabilité entre les réseaux structurés ; après cela vient l'approche des proxy qui permet la conversion des requêtes venant de l'extérieur via des messages web services vers JXTA ; mais JXTA ne s'impose pas encore comme standard ; cela rend cette approche inefficaces pour les réseaux qui sont pas basées sur JXTA ; l'intérêt major de cette approche est la prise en charge de l'aspect sémantique pour la description des ressources via la réécriture des requêtes en prenant en compte l' aspect sémantique ;

Après cette approche on favorise l'approche peerBus qui définit un interlogiciel qui fournit des APIs standard permettant d'assurer le routage des messages et l'accès aux états de routage ; l'inconvénient de cette approche est l'inexistence de l'implémentation et le manque de détails de synergie (scénario) d'utilisation de cette approche ; son avantage c'est la prise en charge de toutes les architectures des réseaux p2p ;

Enfin vient en dernier l'approche Gateway ; qui permet à l'aide de conversion biunivoque d'assurer l'interopérabilité entre deux réseaux p2p ; cette méthode devient trop lourd lors de l'accroissement des réseaux p2p à rendre interopérables.

Concernant l'approche JXTA et d'après nos recherches on a trouvé aucune étude qui expose l'interopérabilité réalisé entre des applications p2p à base de JXTA.

VII. CONCLUSION

Dans cet article, nous avons pu constater qu'il y a peu d'approches pour assurer l'interopérabilité entre les réseaux p2p (à notre connaissance on dénombre six approches), en plus il n'existe pas des évaluations de ces approches. Pour accomplir cette tâche nous avons procéder à l'étude et l'analyse de ces approches, ensuite nous avons procéder à la comparaison de ces approches. Comme perspective a cette étude on propose de concevoir et de réaliser des simulateurs qui permettent de tester ces différentes approches d'interopérabilités ce qui n'existe pas actuellement

COMPARAISON ENTRE APPROCHES D'INTEROPERABILITE DE RESEAUX P2P

	Type d'architecture	Type de service	Méthodes de recherche	nombre de messages échangés	Type de téléchargement	Interopérabilité avec d'autre système	Implémentation
Approche Gateway	Centralisée et décentralisée	Partage de fichiers	- Les index centralisés. - Modèle d'inondation de requêtes	$N*(N-1)$	téléchargement source unique	Seulement avec système p2p	-Difficile et risque d'explosion combinatoire. - L'implémentation n'existe pas.
Approche PeerBus	Tous les types d'architectures	Partage de fichiers	- Les index centralisés. - Modèle d'inondation de requêtes. - Modèle de routage de documents(DHT).	$2*N$	téléchargement source unique	Seulement avec système p2p	- Facile (APIs communes). - L'implémentation n'existe pas.
Approche Proxy	Architecture basé sur JXTA	Partage de fichiers	Recherche à l'aide de service web.	$2*N$	téléchargement source unique	Interopérable avec systèmes p2p et avec les autres systèmes (serveurs	- Basée sur l'utilisation de service Web. - L'implémentation existe (assure l'interopérabilité

						web, serveurs ftp,...)	entre réseau p2p à base de JXTA et serveur Web.
Approche Index	Tous les types d'architectures	Partage de fichiers	Modèle de routage de documents(DHT).	Il n'y a pas de conversion	téléchargement segmenté multi-sources	Interopérable avec systèmes p2p et avec les autres systèmes (serveurs web, serveurs ftp,...)	- L'implémentation existe, mais assure l'interopérabilité entre des applications non p2p ; tel que serveurs http, ftp ou autres.
Approche H-p2pSIP	Réseaux structurés	Partage de fichiers, M.I	Modèle de routage de documents(DHT).	Il n'y a pas de conversion	Non indiqué	Seulement avec système p2p	- L'implémentation existe sous forme de simulation.
Approche JXTA	Architecture basé sur JXTA	Partage de fichiers, M.I	Recherche via les rendez vous (fichier descripteur de ressource en XML)	Il n'y a pas de conversion	Non indiqué	Seulement avec système p2p à base de JXTA	L'implémentation n'existe pas.

REFERENCES

- [1] Standards Information Network. IEEE 100, the Authoritative Dictionary of IEEE Standards Terms, Seventh Edition. New York, NY, IEEE, 2000.
- [2] Standard Computer Dictionary- A Compilation of IEEE Standard Computer Glossaries. NY. 610-1990. ISBN: 1559370793, 1990.
- [3] ISO/IEC 23821. Information technology Vocabulary. Part 1: Fundamental terms. ISO/IEC, Geneva, Switzerland, 1993.
- [4] NATO Allied Data Publication 34 (ADatP-34). NATO C3 Technical Architecture (NC3TA), Version 4.0, 2003.
- [5] Siu Man Lui et Sai Ho Kwok. Interoperability of Peer-To-Peer File Sharing Protocols. ACM SIGecom Exchanges, Vol. 3, No. 3, August 2002, Pages 25 -33.
- [6] Xu, L., Zhou, S., Zhao, K., Qian, W., Zhou, A. PeerBus: A Middleware Framework towards Interoperability among P2P Data Sharing Systems. In: Li, M., Sun, X-H., Deng, Q., Ni, J. (eds.): Proceeding of the 2nd International Workshop in Grid and Cooperative Computing (GCC 2003) (Shanghai, China, December 7-10, 2003), Lecture Notes in Computer Science, vol. 3032. Springer-Verlag, Berlin Heidelberg (2003), pp. 277-284.
- [7] Daniel Olmedilla, et Matthias Palmer. Interoperability for Peer-to-Peer Networks: Opening P2P to the rest of the World. May 2005.
- [8] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmer, and T. Risch. Edutella: A P2P networking infrastructure based on RDF. jun 2002.
- [9] M. Nilsson and W. Siberski. RDF query exchange language (QEL) - concepts, semantics and RDF syntax.2003.
- [10] Evandro S. Rezende, Evgueni Dodonov, Roberta S. Ulson, Marcos A. Cavenaghi, Renata S. Lobato, "Towards Interoperability in P2P World: An Indexing Middleware for Multi-protocol Peer-to-Peer Data Sharing," ICIW, pp.102-105, 2009 Fourth International Conference on Internet and Web Applications and Services, 2009.
- [11] Isaias Martinez-Yelmo; Design and evaluation of interconnecting structured peer-to-peer networks with a Hierarchical topology; university CARLOS III de Madrid; Mars 2010.
- [12] Community, G. : (Gnutella community website) <http://ivre/.gnutella.com/>!
- [13] eDonkey : (edonkey 2000) <http://www.edonkey2000.com/>.
- [14] Kazaa media desktop.
- [15] Institute, S. : (SetiOhome) <http://ivre/.seti.org/science/setiathome.html>.
- [16] Foundation, J.S. : (Jabber) <http://www.jabber.org/>.
- [17] Microsoft : (Msn messenger) <http://me s senger . fr .msn.be/>.
- [18] ICQ : (Icq) <http://web.icqcom/>.
- [19] Dabek, F., Kaashoek, M.F., Karger, D., Morris, R., Stoica, I.: Wide-area cooperative storage with CFS. In: Proceedings of the 18th ACM Symposium on Operating Systems Principles (SOSP '01), Chateau Lake Louise, Banff, Canada (2001)
- [20] Rowstron, A.I.T., Druschel, P. : Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility. In : Symposium on Operating Systems Principles. (2001) 188-201
- [21] Kubiatowicz, J., Bindel, D., Chen, Y., Eaton, P., Geels, D., Gummadi, R., Rhea, S., Weatherspoon, H., Weimer, W., Wells, C., Zhao, B.: Oceanstore : An architecture for global-scale persistent storage. In: Proceedings of ACM ASPLOS. (2000)
- [22] Muthitacharoen, A., Morris, R., Gil, T.M., Chen, B. : Ivy : A read/write peer to peer file system. In: Proceedings of 5th Symposium on Operating Systems Design and Implementation. (2002)
- [23] Clarke, I., Hong, T.W., Miller, S.G., Sandberg, O., Wiley, B. : Protecting free expression online with freenet. IEEE Internet Computing 6 (2002) 40-49
- [24] Quentin Dallons; JXTA : Un Framework Peer-to-Peer Open Source; qdallons@info.fundp.ac.be; Institut d'informatique FUNDP; Namur, Belgique ; 2003.

Intégration d'un Module Morpho-lexical à partir du système Nooj dans une Plateforme d'Analyse de Textes en Arabe standard

Mourad LOUKAM

Natural Language Processing
Team, LMA Laboratory, Faculty of
Sciences, Hassiba Benbouali
University of Chlef, Algeria
mourad.loukam@univ-chlef.dz

Amar BALLA

LMCS Laboratory,
High School of Computer Science,
Algiers, Algeria

Mohamed Tayeb LASKRI

LRI Laboratory, Faculty of
Sciences, Badji Mokhtar
University of Annaba, Algeria

Résumé— Le traitement automatique de la langue Arabe pose de multiples défis à la Recherche en raison des nombreuses spécificités de la langue et la complexité de sa structure. Nous travaillons sur le développement d'une plateforme d'analyse de textes en Arabe standard, en utilisant le formalisme HPSG. L'intégration de ressources libres existantes fait partie de notre stratégie de développement. L'objet de cet article est de décrire une expérience d'intégrer un module morpho-lexical à partir du système open-source Nooj, développé à l'Université Franche Comté (France). Nous décrivons l'apport de ce travail, son originalité et les défis auxquels nous avons été confrontés.

Keywords—Analyse de textes, Arabe standard, HPSG, NOOJ, Ressources linguistiques, TALN

I. INTRODUCTION

Le traitement automatique de la langue Arabe connaît ces dernières années un intérêt important. Ainsi, beaucoup d'applications ont été initiées ou développées pour cette langue, comme : la traduction automatique, la recherche d'information, l'extraction de connaissances, la synthèse de textes, la classification de textes, l'identification thématique, la fouille d'opinions, ... etc. Pour être performantes, toutes ces applications doivent relever de multiples défis posés par les phénomènes complexes inhérents à la structure de la langue Arabe elle-même et ses nombreuses spécificités [1]. Par ailleurs, Ces applications sont confrontées à un problème crucial : c'est celui du manque de ressources diverses, notamment les outils logiciels d'analyse des différents éléments d'un texte, qui doivent être mis en évidence pour être utilisés par ces applications.

Nous contribuons aux efforts de doter la langue Arabe de ressources logicielles en travaillant sur le développement d'une plateforme d'analyse de textes en Arabe standard, en utilisant le formalisme des grammaires syntagmatiques guidées par la tête (Head driven Phrase Structure Grammar, HPSG) [2]. L'objet de cette plateforme est d'offrir un environnement intégré permettant d'avoir toutes les ressources nécessaires à la chaîne d'analyse d'un texte Arabe.

Dans cet article, nous présentons une expérience d'intégration d'un module d'analyse morpho-lexicale fait à partir du système open-source Nooj, développé à l'Université de Franche Comté (France). L'originalité du travail réside dans l'utilisation des fonctionnalités de Nooj, qui n'a pas été fait spécialement pour les grammaires d'unification comme HPSG, dans le processus d'analyse complet de notre

plateforme. Nous verrons que cette intégration permet d'améliorer sensiblement le processus global d'analyse. En revanche, elle pose plusieurs défis à relever, comme le problème d'interfaçage entre les deux systèmes ainsi que l'hétérogénéité des formats de données.

II. MOTIVATIONS

Nous travaillons sur le développement d'une plateforme d'analyse de textes Arabes, en se basant sur le formalisme des grammaires syntagmatiques dirigées par la tête (Head driven Phrase Structure Grammar, HPSG) [3]. En effet, nous pensons que le traitement automatique de l'Arabe ne doit pas viser uniquement le développement d'un "amas" d'applications logicielles, mais doit se faire dans le cadre d'une théorie linguistique qui en assure le cadre formel.

Notre choix s'est porté sur le formalisme HPSG en raison des multiples avantages qu'il offre notamment :

- HPSG hérite de plusieurs formalismes théoriques antérieurs (GPSG, TAG, LFG...) tout en profitant de leurs avantages et en prenant compte leurs insuffisances.

- HPSG opte pour la richesse de la représentation des unités linguistiques : les informations morphologiques, syntaxiques et même sémantiques se retrouvent ensemble dans une même structure, appelée la structure de traits (features structure).

- Pour HPSG, le processus d'analyse se base sur l'unification.

- HPSG utilise un ensemble réduit des règles (schémas) pouvant être appliquées, en principe, à toutes les langues.

- HPSG semble bien adapté au traitement informatique, dont il reprend directement certaines idées (héritage, typage).

L'analyse d'un texte sous HPSG, consiste à « trouver » la tête (l'élément recteur ou dominant) qui orientera ensuite l'analyse vers le dépistage des autres éléments. Bien souvent, il s'agit évidemment d'un véritable défi pour la recherche, notamment pour les phrases contenant des structures plus ou moins complexes (passives, interrogatives, relatives, coordination, ...) et dont l'analyse est réputée difficile.

III. TRAVAUX CONNEXES

Les travaux sous HPSG peuvent être classés en deux catégories : ceux relevant du domaine de la modélisation et ceux se focalisant plutôt sur l'implémentation.

La modélisation vise à proposer une analyse d'un phénomène linguistique donné (phrase passive, interrogative, relative, coordination, ...etc.) en formalisme HPSG. On trouvera dans les proceedings du symposium annuel sur HPSG (<http://hpsg.stanford.edu/>) l'ensemble des travaux de cette catégorie, proposant des analyses d'une multitude de phénomènes linguistiques pour plusieurs langues (par exemple voir [4-5]).

L'implémentation a plutôt pour objectif la mise en œuvre d'outils et d'applications visant à produire des analyses automatiques conformes aux concepts de HPSG dans le but de les tester et de les valider. On peut classer dans cette catégorie, les outils suivants :

- LKB (Linguistic Knowledge Building): est un système de développement grammatical créé par Ann Copestake et son équipe à l'université de Cambridge ([6]). Cet outil n'a pas été conçu spécialement pour les grammaires HPSG, mais il s'agit d'une plateforme de développement pour l'implémentation de toute une famille de grammaires d'unification typées et structurées en traits.

- TRALE : est une plateforme d'implémentation de grammaires HPSG, issue du projet MiLCA et développée à l'Université de Breme (Allemagne). Elle a été utilisée pour abriter CoreGramm un projet visant à contenir un noyau grammatical composé d'un ensemble de langues (Allemand, Anglais, Persan, ..).

- Matrix : une plateforme expérimentale, soutenue par près d'une quinzaine laboratoires de recherches, pour le développement rapide de nouvelles grammaires inspirée des premières grammaires LKB (anglais, japonais, allemand). Il s'agit d'un noyau grammatical universel proposant une signature de base (types généraux, types lexicaux simples, règles de combinaison) et un ensemble de modules paramétrés (questions, négation, coordination, etc.) qui permettent alors de « générer » une analyse sous forme de grammaire typée.

- Enju : un analyseur syntaxique HPSG pour l'anglais, développé au Tsujii laboratory de l'Université de Tokyo. Il est doté d'une grammaire de large couverture qui a été testée sur des modèles probabilistes, notamment dans le domaine biomédical. ([7]).

En ce qui concerne le traitement de l'arabe standard, parmi les rares travaux qui prennent comme cadre de travail le formalisme HPSG, nous pouvons citer le système Maspar ([8]) développé à l'université de Sfax.

IV. LE PROJET PHARAS

Le projet PHARAS (Plateforme d'analyse basée sur le formalisme Hpsg pour l'analyse de l'ARABe Standard) a pour objectif de développer une plateforme intégrée offrant toute la chaîne de traitement d'un texte arabe en vue d'obtenir son analyse selon le formalisme HPSG. L'originalité du projet est de proposer une plateforme d'analyse conçue sur la base d'une intégration des principaux modules concernés par la chaîne d'analyse d'un texte arabe. Et ce contrairement aux outils cités précédemment qui se focalisent sur un module donné de l'analyse et font abstraction des autres. L'objectif premier de la plateforme est de produire une analyse en format HPSG d'un texte en arabe standard.

Depuis le départ, nous avons privilégié l'ouverture et l'orientation "open-source" pour la conception de la

plateforme. L'ouverture signifie faire en sorte, autant que possible, que les différentes ressources du système (paramètres, système de signes, de règles, dictionnaires ...), soient ajoutées ou modifiées pour permettre le développement d'applications pour l'arabe sollicitant tout ou partie de la chaîne d'analyse de la plateforme.

A. 4.1 Fonctionnement général

Un texte en arabe standard introduit sur PHARAS passe par une série de phases de traitement (voir figure 1).

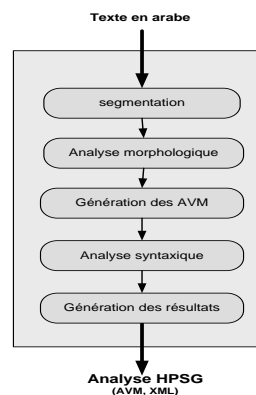


Figure 1. Fonctionnement général de PHARAS

Nous pouvons les résumer ainsi :

1. Phase de segmentation du texte : le texte est décomposé en « mots ».
2. Phase d'analyse morphologique : Après avoir segmenté le texte, on soumet chacun des items rencontrés à une analyse morfo-lexicale.
3. Phase de génération des matrices attribut / valeur HPSG : Cette phase est réalisée par l'analyseur morfo-lexical. Elle consiste à générer, pour chaque item, sa structure de traits sous la forme d'une matrice attributs-valeurs (AVM).
4. Phase d'analyse syntaxique: L'analyse syntaxique en HPSG se base principalement sur l'application du processus d'unification. Il opère sur des structures de traits (AVM) des entrées lexicales des différents mots, déjà générées lors de la phase précédente, ainsi que sur les règles syntaxiques (schémas).
5. Phase de production des résultats : il s'agit de présenter sous forme concrète (AVM ou XML) la représentation syntaxique et sémantique du texte analysé.

B. 4.2 Architecture générale

L'architecture de PHARAS repose sur l'interconnexion de plusieurs sous-systèmes faisant intervenir des ressources et des outils divers (voir figure 2). Nous faisons ci-après une description de cette architecture.

Le sous-système d'analyse morfo-lexicale : après un prétraitement appliqué au texte d'entrée (segmentation), ce système réalise l'analyse morfo-lexicale des éléments du texte.

Le sous-système « signes et règles HPSG » : Ce sous-système est représenté par la signature HPSG retenue, la hiérarchie de types ainsi que les règles à appliquer. Il est composé de trois fichiers : le fichier « Types », le fichier « Règles » et le fichier « Lexique ».

Dans le fichier « Types », on définit la hiérarchie des types utilisés pour décrire les traits. Rappelons que cette

hiérarchie joue un rôle primordial en HPSG puisqu'elle constitue elle-même un ensemble de contraintes sur les structures de traits.

Dans le fichier « Lexique », on stocke toutes les entrées lexicales (verbes, noms, adjectifs, particules) déjà rencontrées ou analysées.

Dans ses premières versions, l'analyse morpho-lexicale était réalisée par un module utilisant un système expert dont les règles codaient les règles morphologiques de la langue Arabe [9] (figure 2). Les avantages de cette méthode étaient l'ouverture et la modularité : il était possible à tout moment d'ajouter ou de modifier une règle morphologique sans

modifier le code du système expert. Mais l'inconvénient était le nombre important de règles (plusieurs centaines) pour prendre en charge tous les phénomènes morphologiques de la langue Arabe.

Dans cet article, nous présentons une expérience d'intégration d'un module morpho-lexical à partir du système libre Nooj (figure 3). Le but recherché est la réutilisation du logiciel et ses fonctionnalités en rapport avec la morphologie de la langue Arabe.

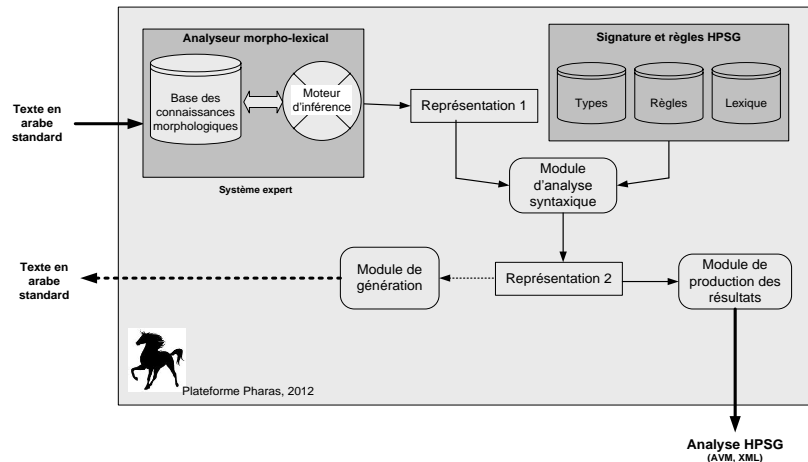


Figure 2. Premières versions de PHARAS : L'analyse morpho-lexicale était faite par un système expert.

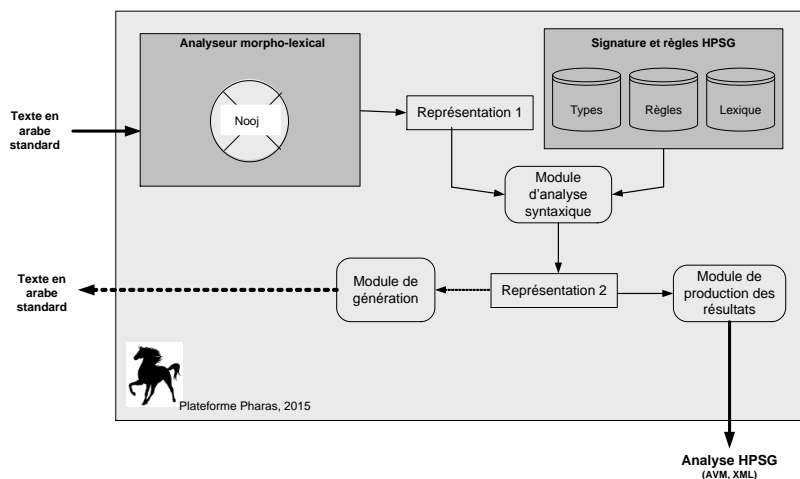


Figure 3. Objet de l'actuelle expérience : L'analyse morpho-lexicale est faite avec un module fait à partir de Nooj

C. 4.3 Introduction au système Nooj

NooJ est un environnement de développement linguistique open-source, développé à l'Université de Franche Comté en France (www.nooj4nlp.net). Il permet de construire des ressources pour des langues naturelles, sous forme de dictionnaires et de grammaires électroniques, et les utiliser dans des applications du TALN [10].

Nooj est utilisé actuellement pour plus d'une vingtaine de langues, dont l'Arabe. Une Conférence internationale annuelle lui est réservée pour présenter des travaux d'utilisation de ce système pour les différentes langues [11-19].

Actuellement, pour la langue Arabe, les ressources existantes sont les suivantes (source www.nooj4nlp.net):

- Noms communs : plus de 19.000

- Verbes : plus de 10.000
- Adjectifs : plus de 5.000
- Particules : plus de 1.200
- Noms propres : plus de 10.000

Parmi les principaux modules existant dans Nooj, nous pouvons citer :

- Création de ressources lexicales : le logiciel permet de créer des dictionnaires spécifiques pour chaque langue contenant les éléments de la langue comme les formes fléchies ou dérivées et devant être utilisées dans le processus d'analyse. La création se fait à partir d'automates en suivant des paradigmes flexionnels et dérivationnels.

- Analyse morphologique et syntaxique d'un texte : le logiciel permet de lancer l'analyse morpho-lexicale d'un

texte. Pour notre expérience, nous ne sollicitons que l'analyseur morpho-lexical de Nooj.

- Traitement de corpus : le logiciel permet de construire des corpus en vue de leur étude.

- Interrogation de textes : le logiciel permet la localisation et la recherche d'éléments dans un texte en utilisant des grammaires restreintes ou des expressions régulières.

V. MODELISATION DE LA MORPHOLOGIE DE LA LANGUE ARABE

Le traitement morpho-lexical dans le système Nooj nécessite la construction d'un dictionnaire contenant tous les lemmes (éléments) de la langue : lemme de base sous la forme d'un nom ou d'un verbe, formes fléchies, formes dérivées, ...etc. Dans ce qui suit, nous décrivons sommairement la formalisation des éléments du processus d'analyse morphologique.

A. 5.1 Construction du dictionnaire des lemmes

La construction du dictionnaire des lemmes, qui constitue le pilier de l'analyse morphologique, se décline en 4 étapes essentielles (figure 4) :

Etape 1 : Etablir le fichier de définition des propriétés (.def). Ce fichier est le premier composant à mettre en œuvre lors de l'étape de construction du dictionnaire. Il contient les différents codages attribués à l'ensemble des catégories grammaticales et à l'ensemble des traits morphologiques correspondant au lexique de la langue Arabe.

Etape 2 : Etablir les grammaires de flexion et de dérivation qui vont permettre de générer les formes fléchies et les formes dérivées à partir des lemmes de base qui sont les verbes et les noms.

Etape 3 : Remplir le fichier de dictionnaire des lemmes de base comme les verbes, les noms. Ces éléments vont servir de point de départ pour générer les autres formes fléchies ou dérivées. Par exemple l'introduction du verbe كَتَبَ/kataba (il a écrit) permettra de générer toutes ses formes fléchies ou dérivées. Par exemple كَتَبَتْ/katabat (elle a écrit) / كَتَبُوا/katabu (ils ont écrit), ...etc.

Etape 4 : Générer le dictionnaire des lemmes proprement dit (on dit aussi compilation) qui génèrent toutes les formes fléchies et dérivées.

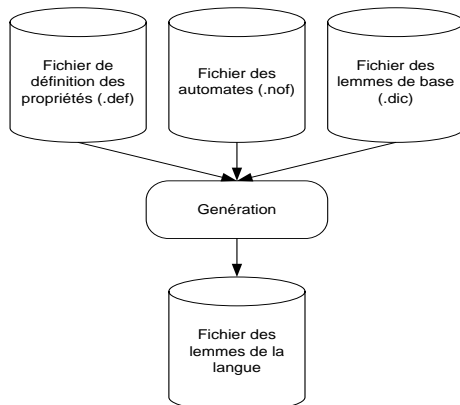


Figure 4 : Génération d'un dictionnaire des lemmes de la langue

B. 5.2 Codage des catégories grammaticales

Nous décrivons ci-après les éléments du lexique devant être pris en compte et codifié pour l'analyse morpho-lexicale. Nous avons retenu le même codage des catégories grammaticales existant actuellement pour la langue Arabe dans le système Nooj [20] .

Catégorie de base : Les catégories grammaticales de base sont au nombre de cinq (Table 1) :

Figure 8. LES CATEGORIES GRAMMATICALES DE BASE

Catégorie	Code	Exemples
Verbes الأفعال	V	كَتَبَ, قَرَأَ
Noms الأسماء	N	كِتَابٌ, مُعَلِّمٌ
Adjectifs الصِّفَات	ADJ	صَغِيرٌ, جَمِيلٌ
Nombres الأعداد	NOMBRE	واحد, إثنان
Noms Propres أسماء العلم	PROP	محمد, عمر

Catégories des Formes dérivées : Elles sont considérées comme des sous catégories des noms dont elles sont dérivées (Table 2) :

Figure 9. LES SOUS-CATEGORIES DERIVEES DES NOMS

Sous-Catégorie	Code	Exemples
Le participe actif إسم الفاعل	PA	حَاكِمٌ, عَامِلٌ
Le participe passif إسم المفعول	PP	مُجْتَبِئٌ, مَكْسُورٌ
Le nom de préférence إسم التفضيل	PREF	أكبر, أصغر
Le nom de lieu/temps ظرف الزمان و المكان	LIEU/TMP	مصنع, موعد
Le nom d'instrument إسم الآلة	INST	مِفْتَاحٌ, مِئْشَارٌ

Catégorie de formes invariables : Correspondent aux pronoms, adverbes, ...etc. (table 3).

Figure 10. LES SOUS-CATEGORIES DES ELEMENTS INVARIABLES

Sous-Catégorie	Code	Exemples
Les pronoms personnels isolés الضمائر المنفصلة	PRON	أنا, أنت, هو
Les pronoms personnels affixés الضمائر المتصلة	PRON+Suff	هَذَا, كَيْفًا, هُمْ
Les pronoms relatifs الأسماء الموصولة	REL	الذي, التي
Les pronoms démonstratifs أسماء الإشارة	DEM	هذا, هؤلاء
Les adverbes أسماء الضرف	ADV	حيثما, كيفما, قبل
Les noms verbaux	NOM-VERB	هيهات, سرعان

Catégorie correspondant aux particules : Correspondent aux particules de la langue comme les prépositions, les adverbes, ...etc (table 4).

Figure 11. LES PARTICULES

<i>Sous-Catégorie</i>	<i>Code</i>	<i>Exemples</i>
Prépositions حروف الجر	PREP	بِ, كِ, مِنْ
Les adverbes حروف الظرف	ADV	حيثُ, فوقَ, أين
Les Conjonctions حروف العطف	CONJ	و, أو, ثُمَّ
Les interrogatives حروف الإستفهام	INTEROG	أ, هل, كم
Les négations حروف النفي	NEGAT	لا, ما, لم, لَنْ
Les explicatives حروف التعليل	EXPLIC	أي
Les vocatives حروف النداء	VOCATIF	يَا, أَيَا
Les particules du futur حروف الإستقبال	FUTUR	س, سوف
Les conditionnels حروف الشرط	COND	إن, لو
Les exceptions حروف الإستثناء	EXCEP	إلا, غير
Les outils de nasb (Verbe) أدوات نصب الفعل المضارع	V-NASB	لَنْ, لِي
Les outils de Jasm أدوات جزم الفعل المضارع	V-JAZM	إِنَّ, لَمْ
Les outils de nasb (Nom) أدوات نصب الاسم	N-NASB	إِنَّ و أخواتها
La particule de définition « ال »	DEF	ال
Prépositions حروف الجر	PREP	بِ, كِ, مِنْ

Catégorie correspondant aux expressions figées : Les expressions figées sont représentées dans le dictionnaire telles qu'elles sont utilisées dans la langue (table 5).

Figure 12. LES EXPRESSIONS FIGÉES

<i>Sous-Catégorie</i>	<i>Code</i>	<i>Exemple</i>
Expressions figées العبارات الثابتة	EXP	رئيس الدولة

C. 5.3 Codage des traits morphologiques

Les résultats attendus de l'analyse morpho-lexicale doivent nous renseigner sur les traits reconnus du lemme analysé (verbe ou nom). Ces traits seront, ensuite, récupérés et traduits en format HPSG, dans la plateforme Pharas. Les tables suivantes décrivent les traits significatifs des verbes et des noms et leur codification. Nous avons retenu la même codification existant actuellement pour la langue Arabe dans Nooj [20].

Figure 13. LES TRAITS DU VERBE

	<i>Trait</i>	<i>Code</i>
Temps	Accompli	Ps
	Inaccompli	Pr
	Impératif	Imp
Mode	Indicatif	R
	Subjonctif	N
	Apocopé	J
Voix	Voix active	A
	Voix passive	P

	<i>Trait</i>	<i>Code</i>
Personne	Premier Personne	1
	Deuxième Personne	2
	Troisième Personne	3
Nombre	Singulier	s
	Dual	d
	Pluriel	p
Genre	Masculin	m
	Féminin	f

Figure 14. LES TRAITS DU NOM

	<i>Trait</i>	<i>Code</i>
Cas	Nominatif	u
	Accusatif	a
	Génitif	i
Tanwin	Nominatif	un
	Accusatif	an
	Génitif	in
Nombre	Singulier	s
	Dual	d
	Pluriel	p
Genre	Masculin	m
	Féminin	f

À ces traits morphologiques peuvent s'ajouter d'autres traits optionnels, portant des informations syntaxiques ou sémantiques, comme [20] :

Figure 15. CODIFICATION DES TRAITS SYNTAXIQUES

<i>Trait</i>	<i>Code</i>
Transitif	Tr
Transitif indirect	TrInd
Intransitif	InTr

D. 5.4 Les opérateurs de transformations morphologiques

NooJ offre des opérateurs génériques de transformations morphologiques permettant la construction d'un lemme quelconque à partir d'un lemme de base, et plus généralement de construire toutes les formes fléchies d'une racine verbale. Ces opérateurs sont :

- <L>: déplacement vers la gauche (Left arrow) ;
- <R>: déplacement vers la droite (Right arrow) ;
- : suppression du dernier caractère (Backspace) ;
- <S>: suppression du caractère courant (Suppr) ;
- <N> : déplacement vers le mot suivant (Next word form) ;
- <P> : déplacement vers le mot précédent (Previous word form) ;

catégorie grammatical « V : pour les verbes », les informations suivants :

- Une information lié à sa transitivité « +Tr », pour désigner que le verbe est transitif.

- L'expression « +FLX=Kataba-Inacc » : désigne un appel au paradigme flexionnel « Kataba-Inacc ». Ce paradigme permet de générer l'ensemble de formes conjuguées à l'inaccompli avec tous les pronoms personnels, incluant le mode indicatif (المرفوع), subjonctif (المنصوب) et apocopé (المجزوم), la voix active (المبني للمعلوم) et passive (المبني للمجهول).

- « +DRV=IsmFa3il_Fa3ala:IsmFa3il_Masc_FLX » : cette expression désigne un appel au modèle de dérivation « IsmFa3il_Fa3ala » à utiliser pour produire la forme dérivée « إسم الفاعل المُذكر » (participe actif masculin) de l'entrée verbale.

Les entrées nominales : Généralement, toute entrée nominale ressemble à :

مُعَلِّم, N+FLX=Nom_Masc+DRV=Masc_Pluriel_Salem:chakle_MP

L'entrée nominale se compose :

- d'un lemme « ex : مُعَلِّم » ;
- d'un code désignant la catégorie grammaticale « N » ;
- Un appel à un modèle flexionnel « +FLX=Nom_Masc », qui permet de générer les différents formes fléchies selon le nombre (singulier, duel), le mode (déterminé, indéterminé) et le cas (génitif, accusatif ou nominatif) ;
- Un appel à un modèle dérivationnel « +DRV=Masc_Pluriel_Salem », qui permet de générer la forme au pluriel de l'entrée, et associer ce dernier à une règle de flexion « :chakle_MP » qui permet de le fléchir selon le cas (génitif, accusatif ou nominatif).

Les formes invariables : Dans la liste des entrées invariables on peut trouver soit des noms indéclinables, tels que les pronoms relatifs, les pronoms personnels...etc, ou des particules (الحروف). Ces entrées ne font appel à aucun paradigme flexionnel/dérivationnel, et seront listées telles qu'elles apparaissent dans le lexique suivi d'un symbole désignant leur catégorie grammaticale (PRON, REL, PREP...etc.). De plus, des informations flexionnelles peuvent être rattachées. Exemple :

أنا, PRON+1+mf+s+u
نَحْنُ, PRON+1+mf+p+u

الَّذِي, REL+m+s
الَّتِي, REL+f+s
بِ, PREP
كِ, PREP

Les expressions figées et les mots composés : Les expressions figées ou mots composés seront listés tels qu'ils apparaissent dans le lexique, suivis des informations les concernant. Par exemple :

صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ, EXP+RELIG
جِهَاتُ الْكُتُبِوتِر, N+m+s+u+TECH
كِرَّةُ الْقَدَمِ, N+f+s+u+SPORT

VI. INTEGRATION DU MODULE D'ANALYSE MORPHO-LEXICALE

Après avoir donné un aperçu des fonctionnalités du système Nooj pour prendre en charge la morphologie de la langue Arabe, nous décrivons comment l'intégration du module morpho-lexical a été réalisée au niveau de notre plateforme d'analyse de textes Arabes utilisant le formalisme HPSG.

La figure 6 montre l'articulation du sous-système d'analyse morpho-lexical par rapport au reste de la plateforme. Nous décrivons ci-après le processus d'analyse de la plateforme ainsi composée :

1. Le texte en Arabe standard est fourni à l'entrée de la plateforme.
2. L'analyse morpho-lexical est réalisée par un appel au sous-système Nooj.
3. L'analyse morpho-lexical fournit ses résultats, qui peuvent être récupérés dans un fichier d'échange avec la plateforme Pharas.
4. La plateforme opère une traduction (qui correspond à une mise en conformité) des résultats fournis par le module Nooj en structures de traits nécessaires à l'analyse postérieure en HPSG.
5. La plateforme amorce l'analyse syntaxique qui utilise le dispositif nécessaire composé par le système de signature, les règles syntaxiques et le moteur d'analyse syntaxique.
6. Le résultat de l'analyse syntaxique est délivré au module de production des résultats qui affiche la représentation de la structure profonde du texte analysé en format HPSG (AVM ou XML).

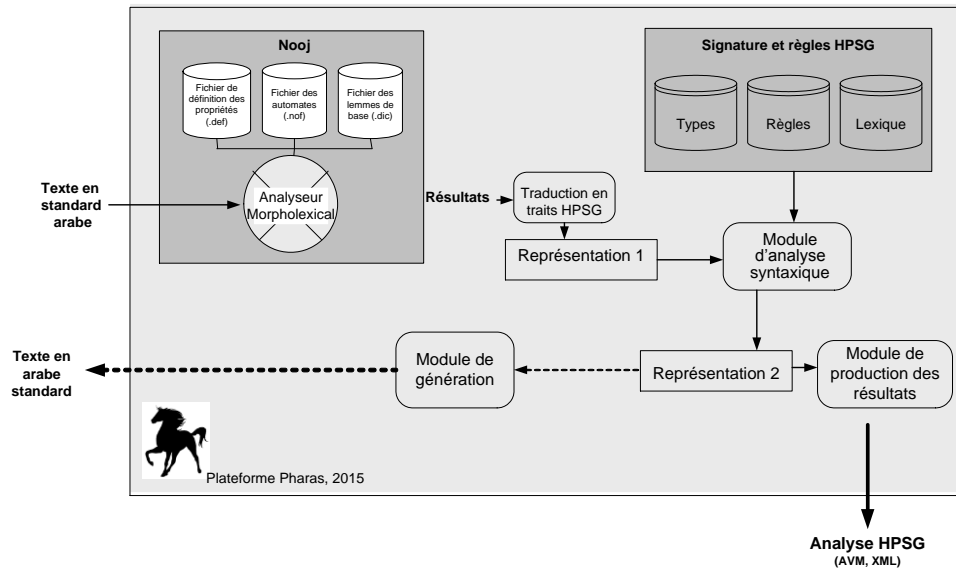


Figure 6. Articulation du module d'analyse morpho-lexicale par rapport au reste de la plateforme

L'interfaçage entre la plateforme d'analyse utilisant le formalisme HPSG et le module d'analyse morpho-lexicale a été réalisée avec une application Java. Elle permet d'actionner le module morpho-lexicale fait à partir de Nooj et de récupérer ses résultats via un fichier d'échange.

La figure 7 donne un aperçu sur l'interface logicielle réalisée entre la plateforme et le module morpho-lexicale. L'interface récupère les résultats de l'analyse morpho-lexicale, dont le nombre de mots analysés, mots inconnus, ...etc.

Parsing ends successfully in:	4s	تأى عملية التحليل بنجاح فى:
Parsing report:		تأى عملية التحليل:
List of used resources (grammars & dictionaries):		القوائم الإلكترونية المستخدمة فى التحليل:
Adjectiv2_Ar.nod; Graph_Morpho_Locked.nod; Nouns2_Ar.nod; Nouns2.nod; Verbes.nod;		
Parsing results:		نتائج التحليل:
Number of parsed words	346	تأى الكلمات
Number of unknown words	80	تأى الكلمات المجهولة
Recognition rate	76%	تأى التحليل

Figure 7 : Application d'interfaçage entre la plateforme d'analyse Pharos et le module morpho-lexicale fait à partir de Nooj

Rappelons que le plus important est de récupérer les traits morphologiques des éléments du texte qui feront l'objet d'une traduction en format HPSG, soit sous forme d'une matrice Attributs-Valeurs (AVM), soit sous la forme d'un fragment XML. La figure 8 donne un exemple de traduction faite des traits obtenus.

Liste of annotated words:				قائمة الكلمات المعروفة
N°	Inflected Form	Lemma	Feature Structure	FS.XML
1	الرحمة	الرحمة	[Form = "الرحمة"; STN = [Catégorie = Nom; Type = Non Proprie; Cas = -; Déterminé = Oui]]	View XML
2	ال	ال	[Form = "ال"; STN = [Catégorie = Particule; Rôle = Définition du Nom]]	View XML
3	رحمة	رحمة	[Form = "رحمة"; STN = [Catégorie = Nom; Type = -; Cas = Génitif; Déterminé = Oui]]	View XML
4	رحمة	رحمة	[Form = "رحمة"; STN = [Catégorie = Nom; Type = -; Cas = Nominatif; Déterminé = Oui]]	View XML
5	رحمة	رحمة	[Form = "رحمة"; STN = [Catégorie = Nom; Type = -; Cas = Accusatif; Déterminé = Oui]]	View XML

Figure 8 Traits morphologiques récupérés de l'analyse morphologique et traduits en format HPSG

VII. EVALUATION

Pour évaluer les performances du système réalisé après intégration, nous avons mené des tests en utilisant un corpus de textes de différents types (religieux, techniques, encyclopédiques, littéraires, articles de journaux, ...etc.).

Les tests réalisés ont montré que le taux de reconnaissance des mots, c'est à dire les mots dont l'analyse morpho-lexicale a pu être faite avec succès, varie entre 80% à 99%, selon les textes.

Nous donnons ci-après deux exemples de textes analysés, le premier est un texte religieux et le second est un texte technique.

Exemple 1 : Texte religieux

Dans ce premier exemple nous avons soumis à l'analyseur un texte religieux (« Hadith حديث » n° 06 du livre « Sahih Al-bukhari صحيح البخاري »). L'analyse morpho-lexicale de ce texte a donné un taux de succès de 85% sur l'ensemble des 466 mots (figure 9). Les cas d'échecs, les 15% restants, s'expliquent par le fait que ce type de texte contient beaucoup de noms propres, qui n'existent pas encore dans les ressources du système.

Parsing results:		ماتج التحليل:	
Number of parsed words	466	عدد الكلمات	
Number of unknown words	67	عدد الكلمات الغير معروفة	
Recognition rate	85%	نسبة نجاح التحليل	

figure9. Exemple d'analyse d'un texte religieux.

Exemple 2 : Texte technique

Dans ce second exemple, nous avons soumis à l'analyse morpho-lexicale un texte technique du domaine d'informatique (extrait de l'encyclopédie Arabe mondiale (الموسوعة العربية العالمية)). Sur les 3428 mots du texte, l'analyse a donné un taux de reconnaissance de 94% (figure 10). Ce qui est appréciable. Les rares cas d'échecs (6%) correspondent aux mots latins transcrits à l'Arabe, comme (سليكون, مايكروسوفت, الترانزستورات...etc.), qui ne figurent pas encore dans les ressources du système d'analyse.

Parsing results:		ماتج التحليل:	
Number of parsed words	3428	مات	
Number of unknown words	184	مات الغير معروفة	
Recognition rate	94%	اج التحليل	

Figure 10 Exemple d'analyse d'un texte technique

VIII. CONCLUSION

Dans cet article, nous avons présenté un travail d'intégration d'un module d'analyse morpho-lexicale open source Nooj, dans notre plateforme d'analyse de textes en Arabe standard basé sur le formalisme HPSG.

Les avantages d'une telle réalisation sont nombreuses, notamment la réutilisation du logiciel qui permet un gain de temps pour le développement des autres modules de la plateforme. Par ailleurs, notre évaluation a permis de montrer que la plateforme peut bénéficier de la bonne couverture linguistique du système d'analyse morpho-lexicale qui est capable d'analyser correctement plus de 90% des mots de la plupart des textes en Arabe standard.

Les défis à surmonter pour réaliser ce travail étaient aussi nombreux. Nous pouvons citer principalement le paramétrage du module pour permettre son intégration dans la plateforme, ainsi que la nécessité de traduire les résultats de l'analyse morpho-lexicale en format conforme au formalisme HPSG.

Pour les développements futurs, nous proposons d'étendre davantage la couverture linguistique du système d'analyse, en ajoutant de nouvelles ressources, comme les entités nommées.

REFERENCES

[1] A.Farghaly and K.Shaalan (2009). Arabic Natural Language Processing : Challenges and Solutions, ACM Transactions on Asian Language Information Processing, Vol. 8, No. 4, pages 14-22.

[2] M.Loukam, A.Balla, M.T.Laskri (2014). An open platform, based on HPSG formalism, for the standard Arabic language . in proceedings of *LREC'2014, Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools*, Reykjavik, Iceland, 27 May 2014, pages 38-42.

[3] C.Pollard & I.A.Sag (1994). Head-driven Phrase Structure Grammar. Chicago: University of Chicago Press and Stanford: *CSLI Publications*.

[4] H.Michael (2011). Null Conjoncts and Bounds Pronouns in Arabic, in *Proceedings of HPSG 2011 Conference*, August 22-25 2011, University of Washington, CSLI Publications.

[5] Hann Michael (2012). Arabic Relativization Patterns: A Unified HPSG Analysis, in *Proceedings of HPSG 2012 Conference*, Chugnam National University of Daejon, South Korea, CSLI Publications, July 18-19 2012.

[6] A.Copestake (2002). Implementing Typed Feature Structure Grammars , *CSLI Publications*, Stanford University, 2002 .

[7] Miyao Y. & Tsujii J. (2005). Probabilistic Disambiguation Models for Wide-Coverage HPSG Parsing , In *Proceedings of ACL-2005*, 2005, p. 83-90 .

[8] Bahou Y., Hadrich Belguith L., Aloulou C., Ben Hamadou A. (2006). Adaptation et implémentation des grammaires HPSG pour l'analyse de textes arabes non voyellés., *Actes du 15e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle RFLA'2006*, Janvier 2006, Tours/France.

[9] M.Loukam, A.Abbache et M.T.Laskri, « Un analyseur morpho-lexical à base de système expert en vue d'une analyse en HPSG », *Actes de la conférence Internationale sur le traitement automatique de la langue arabe CITALA'07*, 18/19 Juin 2007, Rabat/Maroc, p 159-166.

[10] M.Silberstein et A.Tutin, « NooJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE », *Alsic [En ligne]*, Vol. 8, n° 2 | 2005.

[11] K.Bogacki and E.Gwiazdecka . Disambiguating Polish Verbs of Motion. Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2012 International Conference (Paris, France). Cambridge Scholars Publishing, Newcastle., UK: 102-112.

[12] H.Cheikhrouhou. Recognition of Communication Verbs with NooJ. Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrucken, Germany). Cambridge Scholars Publishing, Newcastle., UK: 153-168.

[13] V.Collec-Clerc. Adapting Existing Japanese Linguistic Resources to Build a NooJ Dictionary to Recognise Honorific Forms . Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrucken, Germany).Cambridge Scholars Publishing, Newcastle., UK: 141-152.

[14] T.Declerck and K.Mörth. Porting Persian Lexical Resources to NooJ. Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2012 International Conference (Paris, France). Cambridge Scholars Publishing, Newcastle., UK: 14-23.

[15] K.Dobrovoljc. Introduction to Slovene Language Resources for NooJ. Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrucken, Germany). Cambridge Scholars Publishing, Newcastle., UK: 27-40.

[16] S.Fuentes S. and A.Gupta . Updated Spanish Module for NooJ . Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrucken, Germany). Cambridge Scholars Publishing, Newcastle., UK: 51-62.

[17] Z.Gavriliidou and E.Papadopoulou . Derivation of Multiply Complex Negative Adjectives from Verbal Stems in Greek . Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrucken, Germany). Cambridge Scholars Publishing, Newcastle., UK: 63-68.

[18] I.Keskes, F.Benamara and L.Belghith Hadrich . Discourse Segmentation of Arabic Texts Using Cascade Grammars. Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2012 International Conference (Paris, France). Cambridge Scholars Publishing, Newcastle., UK: 230-232.

[19] D.Najar and S.Mesfar. Political Monitoring and Opinion Mining for Standard Arabic Texts. Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrucken, Germany). Cambridge Scholars Publishing, Newcastle., UK: 115-128.

[20] S.Mesfar. Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en Arabe standard. Thèse de Doctorat. Université de Franche Comté. 2008.

Mobility Models for VANET simulation

Bahidja Boukenadil
STIC Laboratory
Abou Bekr Belkaid University
Tlemcen, Algeria

Mohammed Feham
Dept of Telecommunications
Abou Bekr Belkaid University
Tlemcen, Algeria

Abstract—In the performance evaluation of a protocol for a vehicular ad hoc network, the protocol should be tested under a realistic conditions including, representative data traffic models, and realistic movements of the mobile nodes which are the vehicles (i.e., a mobility model). This paper is a comparative study between two mobility models that are used in the simulations of vehicular networks, i.e., MOVE (MOBility model generator for VEhicular networks) and CityMob, a mobility pattern generator for VANET. We describe several mobility models for VANET simulations. The goal of this paper is to present a number of mobility models in order to offer researchers more informed choices when they are deciding upon a mobility model to use in their performance evaluations. Lastly, we present simulation results that illustrate the importance of choosing a mobility model in the simulation of a protocol in VANET network. Specifically, we illustrate how the performance results of an ad hoc network protocol drastically change as a result of changing the mobility model simulated.

Keywords—VANET, Mobility Model, Simulations, Real World, MOVE, SUMO, CityMob, NS-2 etc.

I. INTRODUCTION

In order to thoroughly simulate a new protocol for an ad hoc network, it is imperative to use a mobility model that accurately represents the mobile nodes (MNs) that will eventually utilize the given protocol. Currently there are two types of mobility models used in the simulation of networks: traces and synthetic models [1]. Traces are those mobility patterns that are observed in real life systems. Traces provide accurate information, especially when they involve a large number of participants and an appropriately long observation period. However, new network environments (e.g. vehicular networks) are not easily modeled if traces have not yet been created. In this type of situation it is necessary to use synthetic models which attempt to realistically represent the behaviors of MNs (i.e., the vehicles nodes) without the use of traces. In this paper, we present several synthetic mobility models that have been proposed for the performance evaluation of vehicular network.

The remainder of the paper is organized as follows. Section II describes related work dedicated to the analysis of the mobility models proposed for VANETs. In section III we illustrate that a mobility model has a large effect on the performance evaluation in simulation of VANET network. Finally, Section IV presents some concluding remarks.

II. MOBILITY MODEL FOR VANET

Since real vehicular traces are not available, a traffic simulator can be used to generate the movement of vehicles.

However, driver behavior on a road is very complex. Driving is interactive, drivers must react to changing road conditions.

Road conditions (e.g., congestion) depend in turn on the drivers' plans and behaviors. Thus, the choice of the traffic simulator in the end influences the relevance and viability of the obtained results.

Vehicular traffic simulators can in general be classified into microscopic and macroscopic simulators. A macroscopic simulator considers such system parameters as traffic density (number of vehicles per km per lane) or traffic flow (number of vehicles per hour crossing some point, usually intersection) to compute road capacity and the distribution of the traffic in the road net. From the macroscopic perspective, vehicular traffic is viewed as a fluid compressible medium, and, therefore, is modeled as a special derivation of the Navier-Stokes equations. In contrast, microscopic simulators determine the movement of each vehicle that participates in the road traffic.

A wide variety of mobility models have been proposed for VANET simulations.

In [2, 3] Saha and Johnson modeled vehicular traffic with a random mobility of nodes over real road topologies extracted from the maps of the US Census Bureau TIGER database. In that work, nodes select one point over the graph as their destination and compute the shortest path to get there. The edges sequence is obtained weighting the cost of traveling on each road on its speed limit and the traffic congestion.

Huang et al. [4] studied taxi behavior. They model the city as a Manhattan style grid with a uniform block size across the simulation area. All streets are assumed to be two-way, with one lane in each direction. Taxi movements are constrained by these lanes. A taxi is characterized by a preferred speed, a maximum acceleration and deceleration, a speed variation associated with the preferred speed at steady state, and a list of preferred destinations, i.e., the taxi stands. The taxis are randomly assigned one of three preferred speeds.

Choffnes et al. [5] designed a street mobility model, named STRAW that incorporates a simple car-following model with traffic control to introduce vehicular congestion, which models real traffic conditions. STRAW relies on street plans to build a road map for the specified target region. It also provides at least one lane in each direction on which vehicles can move. To determine the initial positions of vehicles on the field, it uses a random street placement model that places a vehicle in a lane of a street just before an intersection. If another vehicle is already in that lane, the new vehicle is placed behind the existing one.

Haerri et al. [6, 7] proposed a vehicular mobility simulator for VANETs, called VanetMobiSim, which employs the IntelligentDriver Model (IDM) to determine the speed of vehicles.

In Mahajan et al. [8] three different models were presented: Stop Sign Model (SSM), Probabilistic Traffic SignModel (PTSM) and Traffic Light Model (TLM). The main difference of these models is basically the algorithm used to reproduce stop signs. All roads are modeled as bidirectional roads; the SSM and PTSM assume a single lane in each direction of every road, whereas TLM provides the option for modeling multiple lanes.

Martinez et al [9, 10] present CityMob which is a mobility pattern generator especially designed to investigate different mobility models in VANETs, and their impact on inter-vehicle communication performance. This tool is completely compatible with the ns-2 simulation tool and it was developed in C. CityMob creates urban mobility scenarios and simulates damaged cars using the network to send information to other vehicles, trying to prevent accidents or traffic jams. Figure 1 presents the interface of this model.

Karnadi et al develop a tool MOVE (MObility model generator for VEHicular networks) to which will provide facility for the users to generate real world mobility models for VANET simulations. MOVE [11, 12, 13, 14] tool is built on top of an open source micro-traffic simulator SUMO [15, 16, 17]. The output of MOVE is a mobility trace file that contains information of real-world vehicle movements which can be used by NS-2 [18] or Qualnet [19]. MOVE provides a set of parameters that allows the user to quickly generate real-world simulation scenarios without simulation scripts. Figure 2 shows the interface of this model.

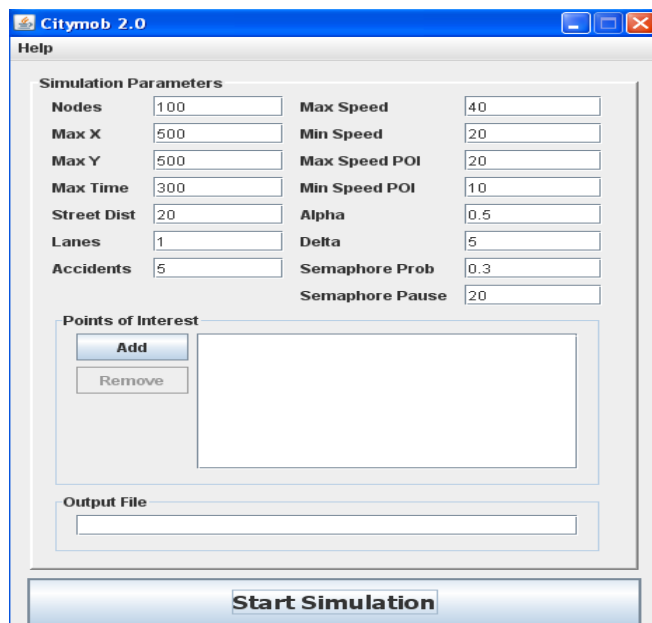


Fig. 1. Interface of CityMob

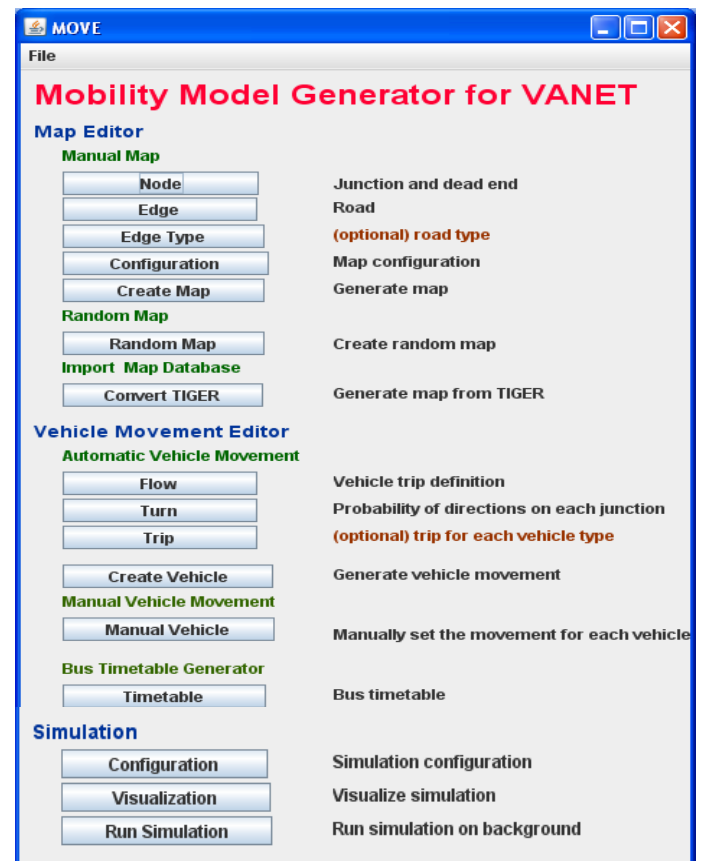


Fig. 2. Interface of MOVE

III. IMPORTANCE OF CHOOSING A MOBILITY MODEL

In this section, we illustrate that the choice of a mobility model can have a significant effect on the performance investigation of an ad hoc network protocol. The results presented illustrate the importance of choosing an appropriate mobility model for the performance evaluation of a given ad hoc network protocol in VANET network. We use ns-2 [34] to compare the performance of the CityMob Model and the MOVE tool via a simulation. The routing of packets is accomplished with the Ad hoc On Demand Distance Vector (AODV) [20]. The parameters for these two mobility models were chosen in a way to simulate path movements that were as similar as possible.

Figure 3 shows an example of road topology for CityMOB, where the darker square area represents the downtown. The vehicles are represented by the Dark rectangles, shadowed rectangles represent vehicles stopped at semaphores, and crosses represent damaged cars sending warning packets.

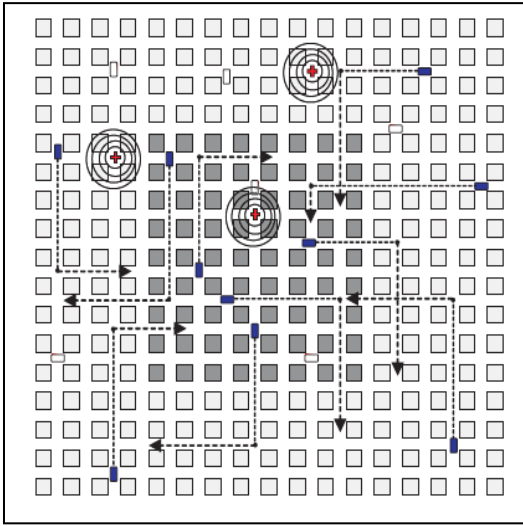


Fig. 3. Downtown scenario of CityMob.

Figure 4 gives an example of a road map created by MOVE for urban scenario, where different roads and junctions with semaphores are represented.

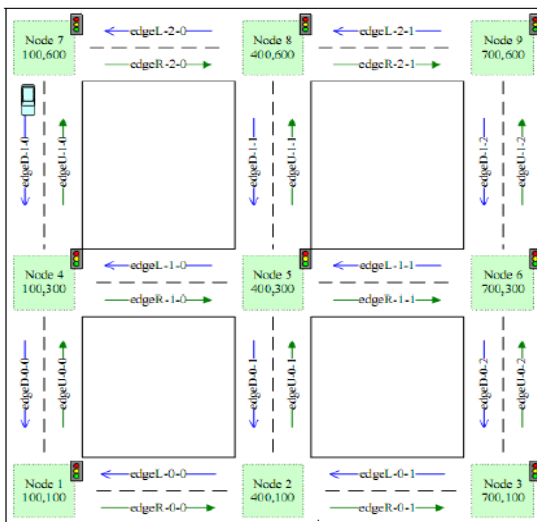


Fig. 4. Road map example of MOVE.

In a source routing protocol, each packet carries the full route (a sequenced list of nodes) that the packet should be able to traverse in its header. In an on demand (or reactive) routing protocol such as AODV, a route to a destination is requested only when there is data to send to that destination, and a route to that destination is unknown or expired. We chose AODV since it performs well in many of the performance evaluations of unicast routing protocols (e.g. [21, 22, 23]).

The ns-2 code used in our simulations of AODV was obtained from [24].

Each simulation run lasted for 300 seconds with a uniform block size of 500 x 500 meters; the maximum speed of vehicles is of 40 m/s.

The number of source nodes from 10 to 50, each of which is a CBR traffic source transmitting UDP packets of a size 512 bytes at the rate of 4 packets per second. All nodes use

802.11 MAC operating at 2Mbps. The propagation model employed in the simulation is TwoRayGround. In our comparison of the two mobility models, we consider the following performance metrics obtained from the AODV protocol: throughput, end-to-end delay and protocol overhead.

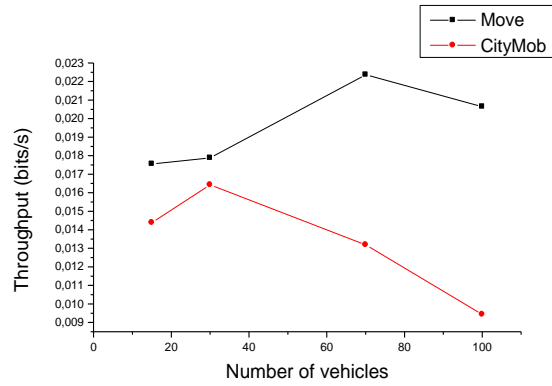


Fig. 5. Throughput vs. number of vehicles.

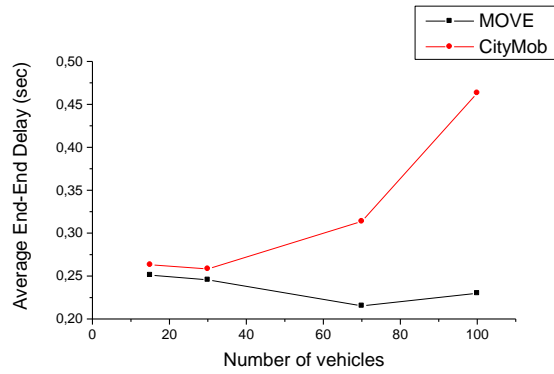


Fig. 6. End-to-end delay vs. number of vehicles.

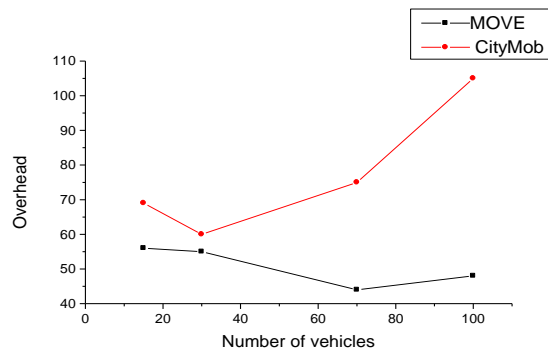


Fig. 7. Overhead vs. number of vehicles.

Figures 5, 6 and 7 illustrate the performance (i.e., Throughput ratio, end-to-end delay and overhead) of AODV with the two mobility models chosen.

The Throughput (in figure 5) of AODV when using MOVE mobility model is higher and more stable than when using CityMob model.

The trace 6 shows that AODV causes a low stable delay with MOVE because the roads are more defined compared to CityMob.

Figure 7 illustrates the overhead AODV required with each of the chosen mobility models. The vehicles moving with CityMob have a higher overhead, as a result this model requires a higher amount of overhead compared to MOVE.

These results confirm the suitability of MOVE tool for simulating VANET.

IV. CONCLUSIONS

In this paper, we compared the performance of two mobility models for VANET Simulation i.e. MOVE (MOBility model generator for VEHicular networks) and CityMob (City Mobility).

Simulation analysis using realistic mobility model for VANET environment show that the performance of the protocol is greatly affected by the mobility model.

The performance of an ad hoc network protocol can vary significantly with different mobility models then, the choice of mobility model in simulating VANET is very important.

The mobility models for VANET should be most closely match the expected real-world scenario. In fact, the anticipated real-world scenario can aid the development of the ad hoc network protocol significantly.

As future work we plan to compare other mobility models discussed above. Results obtained from these studies would certainly facilitate in meeting the challenges associated with future development and evaluation of suitable routing protocols in vehicular networks.

REFERENCES

- [1] M. Sanchez and P. Manzoni. "A java-based ad hoc networks simulator. In Proceedings of the SCS Western Multiconference Web-based Simulation Track, Jan. 1999
- [2] A.K. Saha and D.B. Johnson, "Modeling mobility for vehicular ad hoc networks," in ACM Workshop on Vehicular Ad Hoc Networks (VANET 2004), Philadelphia PA, October 2004.
- [3] U.S. Census Bureau -Topologically Integrated Geographic Encoding and Referencing (TIGER) system <http://www.census.gov/geo/www/tiger>.
- [4] E. Huang, W. Hu, J. Crowcroft, and I. Wassell, "Towards commercial mobile ad hoc network applications: A radio dispatch system," in Sixth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2005), Urbana-Champaign, Illinois, May 2005
- [5] D.R. Choffnes and F.E. Bustamante, "An integrated mobility and traffic model for vehicular wireless networks," in ACM Workshop on Vehicular Ad Hoc Networks (VANET 2005), Cologne, Germany, September 2005.
- [6] J. Haerri, M. Fiore, F. Filali, and C. Bonnet, "Vanetmobisim: generating realistic mobility patterns for vanets," in ACM Workshop on Vehicular Ad Hoc Networks (VANET 2006), Los Angeles, California, September 2006.
- [7] VanetMobiSim. http://en.pudn.com/downloads160/sourcecode/app/detail720214_en.html
- [8] A. Mahajan, N. Potnis, K. Gopalan, and A. Wang, "Evaluation of mobility models for vehicular ad-hoc network simulations," in IEEE International Workshop on Next Generation Wireless Networks (WoNGeN 2006), Bangalore, India, December 2006.
- [9] Francisco J. Martinez, Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni, "CityMob: a mobility model pattern generator for VANETS," in the ICC 2008 workshop proceedings.
- [10] CityMob's source code is available at <http://www.grc.upv.es/>
- [11] MOVE <http://www.cs.unsw.edu.au/klan/move/>.
- [12] F. Karnadi, Z. Mo, K.-C. Lan, "Rapid Generation of Realistic Mobility Models for VANET. Poster Session, 11th Annual International Conference on Mobile Computing and Networking (MobiCom 2005), Cologne, Germany, August 2005.
- [13] F. K. Karnadi, Z. H. Mo, and K. c. Lan, "Rapid generation of realistic mobility models for vanet," in IEEE WCNC, 2007, pp. 2506–2511.
- [14] Karnadi, F.K.; Zhi Hai Mo; Kun-chan Lan; Sch. of Comput. Sci. & Eng., New South Wales Univ., Sydney, NSW, "Rapid Generation of Realistic Mobility Models for VANET". IEEE Xplore
- [15] SUMO Simulation of Urban MObility. <http://sumo.sourceforge.net/>.
- [16] A. Uchiyama, "Mobile Ad-hoc Network Simulator based on Realistic Behavior Model, 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2005), Urbana
- [17] Deutsches Zentrum für Luft-und Raumfahrt e.V. (DLR). Sumo-simulation of urban mobility. <http://sumo.sourceforge.net/>.
- [18] The Network Simulator ns 2. <http://www.isi.edu/nsnam/ns/index.html>.
- [19] QualNet Network Simulator. <http://www.scalable-networks.com/>.
- [20] Perkins, C. and Royer, E. (1999). Ad hoc On Demand Distance Vector (AODV) Routing. In Proceedings 2nd Workshop on Mobile Computing Systems and Applications. New Orleans, LA, USA: IEEE, February 1999, pp. 90–100.
- [21] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva. Multi-hop Wireless ad hoc network routing protocols. In Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), pages 85–97, 1998.
- [22] B. Boukenadil, M. Feham, "Comparison between DSR, AODV and DSDV in VANET using CityMob", in Proc. the 1th International Conference on New Technologies and Communication (ICNTC'2012), Chlef, Algeria, 05-06 December 2012.
- [23] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark. Routing protocols for mobile ad-hoc networks - a comparative performance analysis. In Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), pages 195–206, 1999.
- [24] The Rice University Monarch Project. Monarch extensions to the ns simulator. URL: <http://www.monarch.cs.rice.edu/>. Page accessed on May 30th, 2002

New Half-mode Waveguide Designs based on SIW Technology for X band Applications

M. A. Rabah¹, M. Abri¹, J. W. Tao² and T. H. Vuong²

¹Laboratoire de Télécommunications

Département de Génie Electrique, Faculté de Technologie, Université de Tlemcen
BP 230, Pôle Chetouane, 13000 Tlemcen- Algérie.

²Laboratoire de Laplace, Université de Toulouse-France.

Abstract—In this work a systematic analysis of the signal integrity performance of substrate integrated waveguide structures (SIW) is reported. It is numerically demonstrated how a very high transmission in X-band can be successfully achieved by using this type of structure. This recently technology resume an advantages of a traditional waveguide and coplanar line. In this work we present a parametric study for choosing the best parameter of SIW waveguide within the frequency range [8-12] GHz, this frequency range is specifying for the X-band. A HMSIW (half mode substrate integrated waveguide) design is indicated. It is demonstrated numerically by comparison with simple waveguide using CST Microwave Studio how the transmission in X-band can be successfully achieved with HMSIW.

Keywords-X-band; Substrate Integrated Waveguide (SIW); Half Mode; design; transition.

I. INTRODUCTION

With the growing technical and performance requirement on wireless systems for communication and sensor applications, waveguide components such as couplers, detectors, isolators, phase shifters and slotted lines are commercially available for various standard waveguide bands, this components present properties of low loss and high power, because the fabrication of a transmission line is limited in some of practical applications and it is also difficult to manufacture them in mass production as the rectangular waveguide, the use of this type of guide presents a traditional solution for guidance signal [1].

Conventional waveguide components are typically bulky and expensive; and difficult to integrate with other microwave and millimeter wave planar circuits. On the other hand, microstrip lines are compact and low cost, and have been widely used, but they are relatively provides loss.

Recently, an alternative low cost technology utilized to build a wide range of mm-wave components. The technology is based on using printed circuit boards to imitate waveguides and is called substrate integrated waveguides (SIW), and is known for its numerous advantages such as relatively high Q, ease of integration, compact size, and low cost [2-7] as alternative structures to conventional transmission lines. Substrate integrated waveguides are synthetic rectangular waveguides formed by top and bottom metal layers which embed a dielectric slab and two sidewalls of metallic vias as shown in Fig.1. This technology is one of the most popular and the most developed platforms.

The concept of the half-mode substrate integrated waveguide (HMSIW) was proposed recently in [8]. It can be observed, on one hand, that both the waveguide width and the surface area of the metallic sheets are reduced by nearly half compared with the SIW, and on the other hand, that the fabrication complexity is maintained at the same level as for

the SIW. Several components have been developed based on the HMSIW since its introduction [9-11].

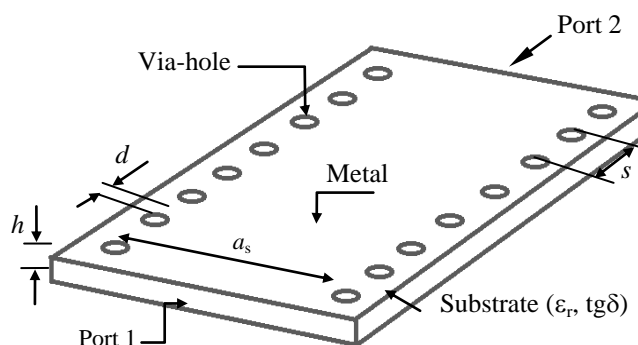


Figure 1. Substrate integrated waveguide (SIW) design present

To going to the best parameters geometrics ; a parametric study of the diameter and distance of via is presented in this paper for a SIX component operating in X-band for TE₁₀ mode.

After giving the geometric parameters of SIW operating in X-band some waveguide miniaturization techniques will be studied in SIW structures. HMSIW (Half Mode Substrate Integrated Waveguide) is presented and a comparison between SIW and HMSIW has been made.

II. GEOMETRIC AND PARAMETRIC STUDY

A. SIW Design Procedure

SIW devices can be thought as a form of dielectric filled waveguide (DFW), therefore the starting point can be DFW. For TE₁₀ mode, the dimension "b" is not important, as it does not affect the cut off frequency of the waveguide. Therefore, the substrate can be at any thickness; it only affects the dielectric loss.

For a rectangular waveguide, cut off frequency of arbitrary mode is found by the following formula:

$$f_c = \frac{c}{2\pi} \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \quad (1)$$

Where:

- c: speed of light
- m, n: mode numbers
- a, b: dimensions of the waveguide

For TE₁₀ mode, the much-simplified version of this formula is:

$$f_c = \frac{c}{2a} \quad (2)$$

For DFW with same cut off frequency, dimension " a_d " is found by:

$$a_d = \frac{a}{\sqrt{\epsilon_r}} \quad (3)$$

Having determined the dimension " a_d " for the DFW, we can now pass to the design equations for SIW [4].

$$a_s = a_d + \frac{d^2}{0.95s} \quad (4)$$

Where:

- d : diameter of the via
- s : pitch (distance between the vias)

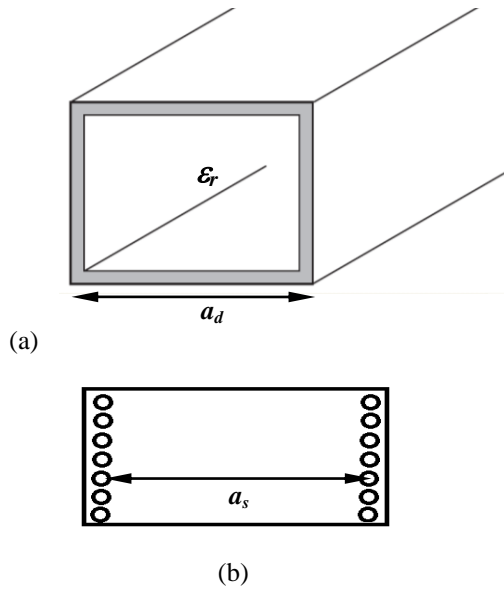


Figure 2. Dimensions for :(a) DFW, (b) SIW

For SIW design, the following two conditions are required[4]:

$$d < \frac{\lambda_g}{5} \quad (5.a)$$

$$s \leq 2d \quad (5.b)$$

Where λ_g (guided wavelength) is [12]:

$$\lambda_g = \frac{2\pi}{\sqrt{(2\pi f)^2 \epsilon_r - \left(\frac{\pi}{a}\right)^2}} \quad (6)$$

Our goal is to produce a SIW waveguide with best transmission, why we must adapt our structure.

A transition from a microstrip line to the waveguide SIW is necessary.

The formula used to calculate the guide impedance of the SIW, which is given by (7) [1]:

$$Z_{pi} = Z_{TE} \frac{\pi^2 h}{8a_s} \quad (7)$$

For the calculation of the guide impedance, it is also necessary to calculate the wave impedance of TE mode, which is given by(8) [1]:

$$Z_{TE} = j\omega \frac{\mu}{\gamma} = \omega \frac{\mu'}{\beta} = \sqrt{\frac{\mu}{\epsilon}} \times \frac{\lambda_g}{\lambda} \quad (8)$$

B. Simulation Results

In order to operate the SIW in the X-band applications we investigated these equations and the geometry presented in Fig.3 are used for testing. The cutoff frequency of TE₁₀ mode in SIW is selected about 7.5 GHz. The waveguide is designed on ArlonIso 917 (lossy) with dielectric constant of $\epsilon_r = 2.2$ and $\tan\delta = 0.0013$ and with the following considerations:

- TE₂₀ mode should not be excited in SIW.
- The physical width should be as wide as possible so that ferrite slabs can be inserted.

Table 1. The Geometrical Parameters of SIW topology.

Width	a_s	15.8mm
Effective width	a_d	13.48mm
Via holes diameter	d	3.3mm
Distance between vias	s	4.95mm
Substrate thickness	h	0.508 mm
50 ohm Microstrip Line width	W_1	1.3mm
18..23ohm Microstrip Line width	W_2	6.16mm
Transition length	L	26.84mm

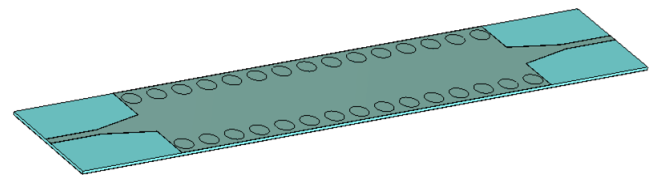


Figure 3. Topology of the Substrate Integrated Waveguide transition.

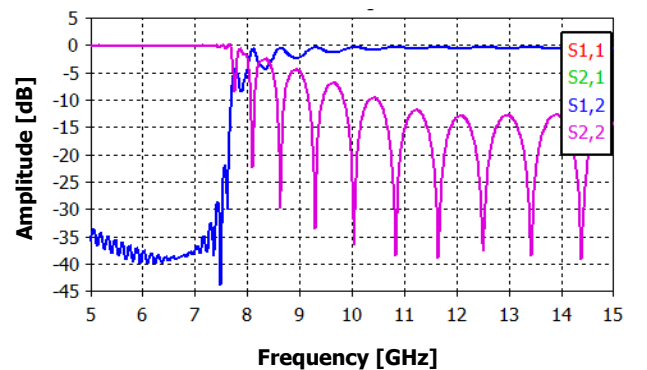


Figure 4. SIW waveguide [S] parameters.

C. Parametric study

In this section, a parametric study is proposed to observe the influence of distance and diameter vias on the transmission coefficient and the input return loss for the X-band waveguide.

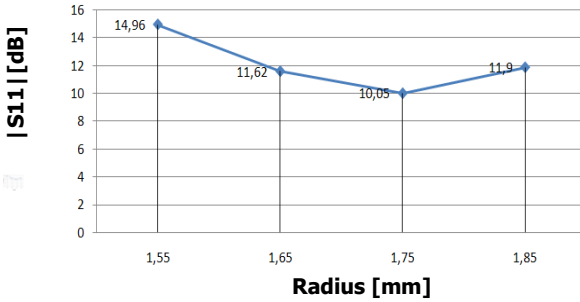


Figure 5. Input return losses for $r=1.55, 1.65, 1.75, 1.85\text{mm}$ and $p=1.5r$ at 9 GHz

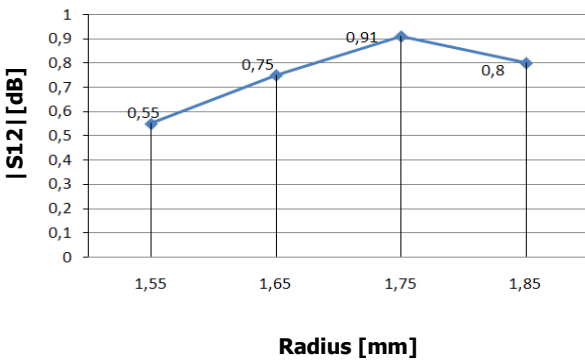


Figure 6. Transmission coefficient for $r=1.55, 1.65, 1.75, 1.85\text{mm}$ and $p=1.5r$ at 9 GHz

As it is shown in Fig. 5 when the radius value is incremented, the reflection coefficient increases which shows the advantage of choosing small values of the radius.

Fig.7 and Fig.8 shows the results of variation of the distance between the vias for the smallest diameter ($r = 1.55\text{ mm}$)

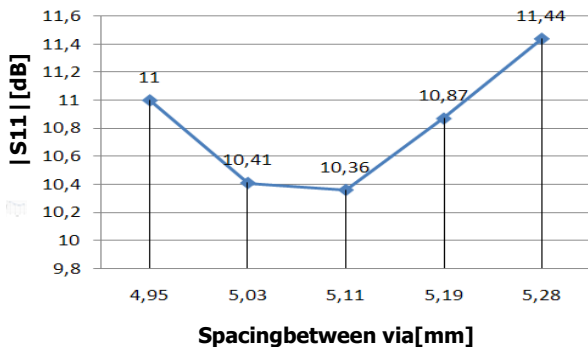


Figure 7. Input return losses for $p=4.95, 5.03, 5.11, 5.19, 5.28\text{ mm}$ at 9GHz.

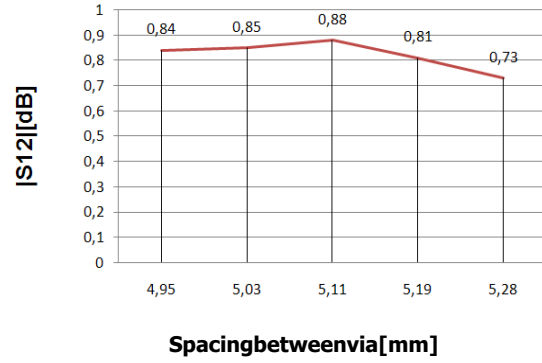


Figure 8. Transmission coefficient for $p=4.95, 5.03, 5.11, 5.19, 5.28\text{ mm}$ at 9GHz.

III. Comparison between SIW and HMSIW

A. HMSIW theory

A miniaturization technique is documented in [12] with the introduction of a half-mode SIW (HMSIW). The half-mode SIW is able to propagate guided waves in only half the width of the standard SIW. The symmetric plane along the direction of transmission is considered a perfect magnetic wall. The guided propagation is unchanged when cutting through the magnetic wall. The HMSIW interconnects require slightly more than half of the SIW width to operate with the same cutoff frequency, as can be observed in Fig.9 where the SIW is truncated at the lower edge of the microstrip. In this configuration, the added width required is equal to half of the microstrip width. The HMSIW is particularly beneficial in that it only requires a single layer PCB fabrication.

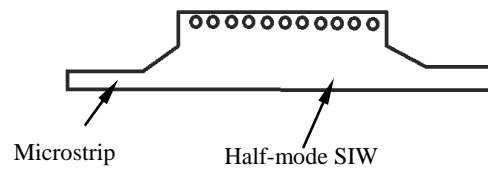


Figure 9. HMSIW interconnect.

B. Design and simulation results

The configuration of the X-band SIW waveguide proposed is shown in Fig. 10, which is composed of 11 holes vias. The half wave guide SIW must be designed on a substrate Arlon ISO 917 (loss) which has a relative permittivity ϵ_r of 2.2, dielectric thickness h of 0.508 mm, a loss tangent of about 0,0013, and the thickness of the conductor of 0.05 mm. The diameter of the vias was 2.3 mm and the distance between the via holes is to be 3.5 mm.

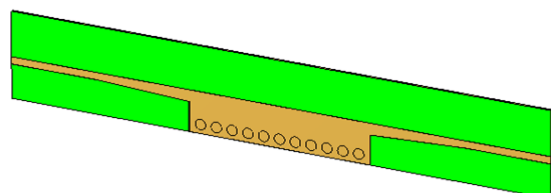


Figure 10. HMSIW interconnect obtained with CST.

Fig. 11 shows the distribution of the electric field in HMSIW operating in X-band.

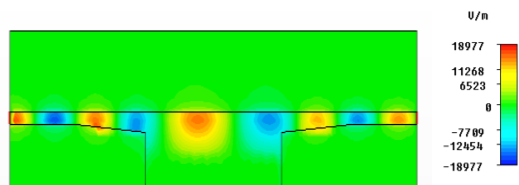


Figure 11. Distribution of electric field in HMSIW

In this figure, it can be seen that the electric field can propagate in only one-half mode SIW.

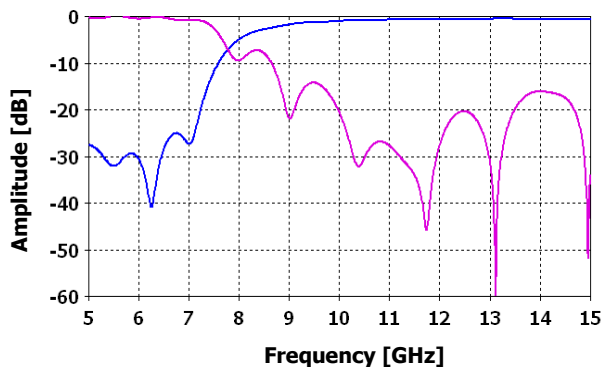


Figure 12. HMSIW waveguide [S] parameters.

In the Fig. 11, the transmission coefficient for HMSIW gives the best result at $f = 10$ GHz (injected frequency of 9 GHz), and we can see from the results of two structures are very approach; and input return loss in Fig. 12 shows good results for the HMSIW structure since in the frequency range [9-12] GHz return loss input is less than -10 dB.

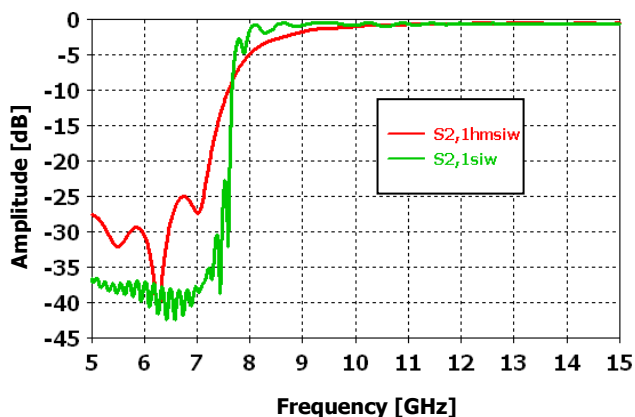


Figure 13. Comparison between the transmission coefficient of SIW and HMSIW obtained by CST Microwave studio.

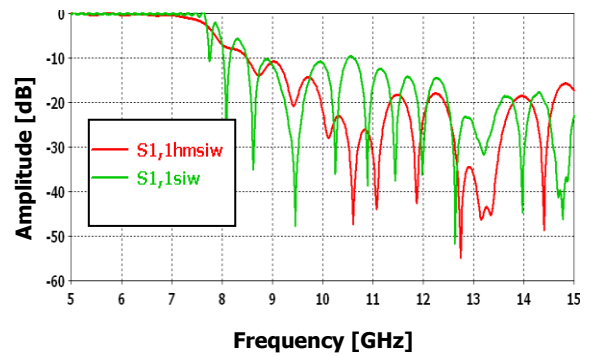


Figure 14. Comparison between the return loss of SIW and HMSIW obtained by CST Microwave studio.

IV. CONCLUSION

In this paper, the SIW waveguide and half-mode SIW (HMSIW) structure with via-hole has been successfully demonstrated. Based on rectangular waveguide theory, specific design rules of SIW for the design of via-hole array and effective width of broad sidewall among others were reviewed. The cutoff frequency of the X-band was deliberately selected to avoid higher-order modes while maximizing the broad side width and a SIW was modeled and simulated out of Arlon Iso 917 (lossy) material using the SIW design rules described in this paper. For transition from microstrip to SIW, guided-wave impedance of SIW was calculated in order to decide the matched width of microstrip, then simulations with a commercial software were followed up to verify the matching condition. In order to reduce size of the structure and conserving all the properties of the SIW another topology is employed which is HMSIW. It was observed that the cutoff frequency is not changed and good results are obtained in terms of return loss and transmission. The circuits with HMSIW can be used for X-Band applications.

REFERENCES

- [1] Yong Ju BAN, "Tunable Ferrite Phase Shifters Using Substrate Integrated Waveguide Technique," thesis, December 2010.
- [2] K. Wu, D. Deslandes, Y. Cassivi, "The substrate integrated circuits – A new concept for high frequency Electronics and Optoelectronics", Microwave Review, nr. 12, December 2003, pp. 2-9.
- [3] Xiao-Ping Chen, Ke Wu, Zhao-Long Li, "Dual-Band and Triple-Band Substrate Integrated Waveguide Filters With Chebyshev and Quasi-Elliptic Responses", IEEE Trans. Microwave Theory Tech., vol. 55, pp. 2569-2577, December 2007
- [4] Y. Cassivi, L. Perregini, P. Arcioni, M. Bressan, K. Wu, and G. Conciauro, "Dispersion characteristics of substrate integrated rectangular waveguide," in IEEE Microwave Wireless Compon. Lett., vol. 12, Sep. 2002, pp. 333-335.
- [5] www.cst.com.
- [6] Y. Cassivi and K. Wu, "Low cost microwave oscillator using substrate integrated waveguide cavity," IEEE Microw. Wireless Compon. Lett., vol. 13, no. 2, pp. 48-50, Feb. 2003.
- [7] J. A. Ruiz-Cruz, M. A. E. Sabbagh, K. A. Zaki, J. M. Rebolgar, and Y. Zhang, "Canonical ridge waveguide filters in LTCC or metallic resonators," IEEE Trans. Microw. Theory Tech., vol. 53, no. 1, pp. 174-182, Jan. 2005.
- [8] W. Hong, B. Liu, Y. Q. Wang, Q. H. Lai, and K. Wu, "Half mode substrate integrated waveguide: A new guided wave structure for microwave and millimeter wave application," in

Proc. Joint 31st Int. Infrared Millimeter Wave Conf./14th Int. Terahertz Electron. Conf., Shanghai, China, Sep. 2006, pp. 18–22.

- [10] B. Liu, W. Hong, Y. Q. Wang, Q. H. Lai, and K. Wu, "Half mode substrate integrated waveguide (HMSIW) 3 dB coupler," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 1, pp. 22–24, Jan. 2007.
- [11] Y. Wang, W. Hong, Y. D. Dong, and B. Liu *et al.*, "Half mode substrate integrated waveguide (HMSIW) bandpass filter," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 4, pp. 256–267, Apr. 2007.
- [12] B. Liu, W. Hong, Y. Zhang, H. J. Tang, X. X. Yin, and K. Wu, "Half mode substrate integrated waveguide 180° 3-dB directional couplers," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 12, pp. 2586–2592, Dec. 2007.
- [13] Y. Cheng, W. Hong, and K. Wu, "Half mode substrate integrated waveguide (HMSIW) directional filter," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 7, pp. 504–506, Jul. 2007.
- [14] J. E. Rayas-Sanchez and V. Gutierrez-Ayala, "A General EM-Based Design Procedure for Single-Layer Substrate Integrated Waveguide Interconnects with Microstrip Transitions", *IEEE MTT-S Int. Microwave Symp. Dig.*, Atlanta, GA, Jun. 2008, pp. 983-986.

Stockage distribué de données dans les réseaux de capteurs sans fil

Amina Chikhaoui, Djamel-Eddine Zegour, Walid-Khaled Hidouci

Ecole Nationale Supérieure d'Informatique (ESI)

Alger, Algérie

{a_chikhaoui, d_zegour, w_hidouci}@esi.dz

Résumé— Les progrès technologiques dans les domaines de la microélectronique, des communications sans fil, couplés aux efforts de miniaturisation et à la réduction des coûts de production des composants électroniques, ont donné naissance à des nouvelles générations de réseaux sans fils. Ces derniers offrent beaucoup d'avantages notamment en termes de déploiement. Parmi ces générations, on trouve les réseaux de capteurs sans fil (WSN). La montée en puissance des capteurs et leur diversité qui ne cessent de croître sont accompagnées de nouvelles applications et problématiques. Le stockage distribué de données est l'un de ces nouveaux challenges. Dans cet article, nous allons proposer une nouvelle méthode de stockage de données pour les WSN basée sur une variante de structures de données distribuées et scalables appelée: PBST* (Distributed Partitionned Binary Search Tree) afin d'optimiser le stockage et la collection de données dans ce type de réseaux.

Keywords—Réseaux de capteurs sans fil; Stockage data centric; GHT ;

I. INTRODUCTION

Concrètement, un WSN est composés d'un ensemble de capteurs intelligents miniaturisés alimentés par des batteries. Ces derniers sont dotés des moyens qui leur permettent de collecter, stocker, traiter et transmettre les données collectées à un centre de collecte via une connexion sans fil [1].

Depuis leur apparition, les réseaux de capteurs sans fil provoquent un intérêt croissant au sein des communautés scientifiques et industrielles. En effet, ils sont appliqués aujourd'hui dans plusieurs domaines d'applications. Nous citons par exemple les domaines : militaire, surveillance, environnemental, médical, domestique, commercial, etc. Quelques exemples des applications des WSN sont décrits dans [2].

Selon les interactions entre le réseau de capteurs et la station de base, nous distinguons trois modèles principaux : modèle de mesure périodique, modèle de détection d'événements et modèle de transmission suite à des requêtes [3].

Dans le premier modèle, tous les capteurs envoient périodiquement leurs mesures à la station de base. Dans le deuxième modèle, les capteurs envoient les mesures seulement lorsqu'il y a un événement qui se produit et dans le troisième modèle, les capteurs mesurent des phénomènes et stockent ces mesures dans leurs mémoires, ils envoient ces dernières seulement lorsqu'ils reçoivent des requêtes de la station de base.

La contrainte principale dans un WSN est le budget d'énergie limité d'un nœud capteur avec l'exigence d'une longue durée de vie du réseau. La consommation d'énergie

devient ainsi leur problème fondamental. Plusieurs protocoles et algorithmes ont été proposés dans la littérature au niveau de toutes les couches de la pile protocolaire du WSN afin de minimiser la consommation de l'énergie.

Dans un réseau de capteurs, les communications coûtent chères par rapport aux traitements locaux des données. Parfois, il vaut mieux stocker et traiter des données localement plutôt que de faire des transmissions. Le modèle de stockage de données peut beaucoup influencer la consommation d'énergie globale du réseau.

Une stratégie de stockage des données mal conçue augmente les frais généraux de communication, dissipe l'énergie utile et réduit la durée de vie des réseaux de capteurs. En revanche, une bonne stratégie de stockage peut considérablement optimiser la consommation d'énergie, réduire les délais de traitement des requêtes et prolonger la durée de vie d'un réseau de capteurs.

Le stockage distribué de données est un nouvel axe de recherche dans les WSN qui consiste à stocker les données détectées dans le réseau lui-même, afin de les récupérer plus tard par des protocoles de récupération appropriés.

L'objectif de cet article est la proposition d'un nouveau protocole de stockage distribué de données pour les WSN. Ce protocole est basé sur PBST* [4], une variante de structures de données distribuées et scalables (SDDS) [5, 6]. La puissance de cette structure réside dans l'utilisation de la mémoire distribuée, le traitement parallèle, les requêtes à intervalle, l'équilibrage de la charge et la capacité de stockage théoriquement illimitée. Elle permet de faire des recherches en temps logarithmique.

La suite de cet article est organisée comme suit : dans la deuxième section nous allons présenter les travaux antérieurs de stockage de données dans les WSN et de SDDS. Dans la troisième section, nous décrirons PBST*. Notre proposition sera présentée dans la quatrième section. Enfin, la cinquième section conclut l'article.

II. TRAVAUX ANTERIEURS

Notre travail tire profit de deux grands axes de recherche : le stockage de données dans les WSN et les structures de données distribuées et scalables. Dans ce qui suit, nous allons présenter les travaux antérieurs de chaque domaine.

A. Stockage de données dans les réseaux de capteurs

Les différents modèles de stockage de données dans les WSN existants sont classés en trois catégories [7] : stockage externe, stockage local et stockage distribué.

Dans Le modèle de stockage externe [8, 9], les capteurs envoient toutes leurs mesures à la station de base et ces mesures sont traitées complètement sur la station de base. Si tous les évènements détectés par les capteurs sont utilisés par la station de base, ce modèle serait le préféré. Mais, dans certaines applications, les requêtes envoyées vers la station de base n'englobent pas l'intégralité des évènements détectés par les capteurs. Ce qui engendre un gaspillage de l'énergie.

Le modèle de stockage local [10, 11] fonctionne de manière différente. Les capteurs stockent les évènements détectés dans leurs mémoire afin d'éviter la transmission des évènements qui n'intéressent pas les utilisateurs. Ils envoient des données seulement lorsqu'ils reçoivent une requête sur les mesures qu'ils possèdent. Le problème dans ce modèle est quand un utilisateur cherche une donnée, il ne sait pas où est-elle stockée. Donc, une requête sera diffusée sur tous les nœuds du réseau. Ce modèle de stockage n'est pas intéressant dans le cas où le nombre de requêtes envoyées par la station de base et le nombre de nœuds sont importants car la diffusion des requêtes dans tout le réseau devient coûteuse.

Le stockage distribué de données consiste à stocker les données dans le réseau lui-même selon une stratégie de stockage. Dans ce modèle, un nœud qui capte une donnée n'est pas forcément le nœud qui stocke cette dernière.

Le stockage data centric (DCS) [7, 12, 13, 14] est un type de stockage distribué de données. Il reprend le paradigme de stockage de données structurées dans les réseaux pair-à-pair. Dans ce modèle, chaque nœud dans le réseau stocke un type de donnée appropriée. De cette façon, les requêtes sur un type d'évènements seront routées directement vers le nœud concerné par ce type. Le premier travail réalisé dans ce domaine est GHT (Geographic Hash Table).

GHT [12] est le premier protocole DCS proposé en 2002 par Ratnasamy et al. La motivation de cette recherche était la gestion efficace de la grande quantité de données détectées par les réseaux de capteurs à grande échelle à l'aide d'un algorithme de dissémination de données évolutif, auto-organisé et efficace en énergie [15].

GHT suppose que chaque capteur connaît ses coordonnées géographiques et que chaque donnée est décrite par une méta-donnée. Ce protocole exploite deux primitives: *Stocker* (d, k) pour le stockage de données et *Récupérer* (k) pour la récupération des données. La primitive *Stocker*(d, k) prend en entrée une donnée d (la donnée captée) et sa métadonnée k . En appliquant une fonction de hachage h sur k , il sélectionne une paire de coordonnées $(x, y) = h(k)$ et il utilise le protocole de routage GPSR [16] pour transmettre un paquet de données à cette emplacement. Le nœud le plus proche de cette position géographique est choisi comme home node. De la même façon, la primitive *Récupérer* (k) applique la même fonction de hachage sur la méta-donnée k pour obtenir les coordonnées (x,y) du home node, et par l'intermédiaire de GPSR, il envoie une requête de récupération au point (x, y) . Lorsque cette demande atteint le home node, ce dernier répond en fournissant les données stockées pour cet évènement.

Cette solution achemine les requêtes de stockage et de récupération directement au home node en évitant l'inondation qui est une opération gourmande en énergie.

Dans [7], cette approche a été démontrée comme une technique de dissémination de données économique en énergie par rapport au stockage externe et stockage local.

Dans la figure 1, les capteurs sont représentés par des cercles bleus. Un nœud producteur (cercle rouge) capte un nouvel évènement. Ce nœud producteur hache cet évènement afin de trouver la position du home node. Ensuite, il transmet l'évènement détecté au home node désigné par un cercle vert. À son tour, un autre nœud consommateur utilise la même fonction de hachage et récupère les données stockées sur le home node.

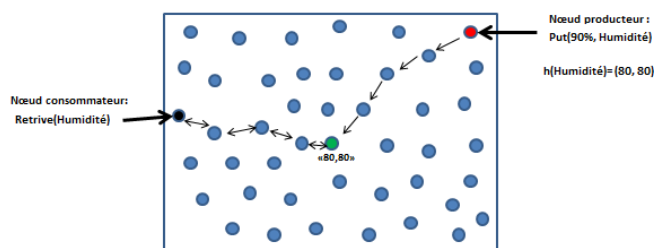


Fig. 1. Diagramme d'exécution des requetes

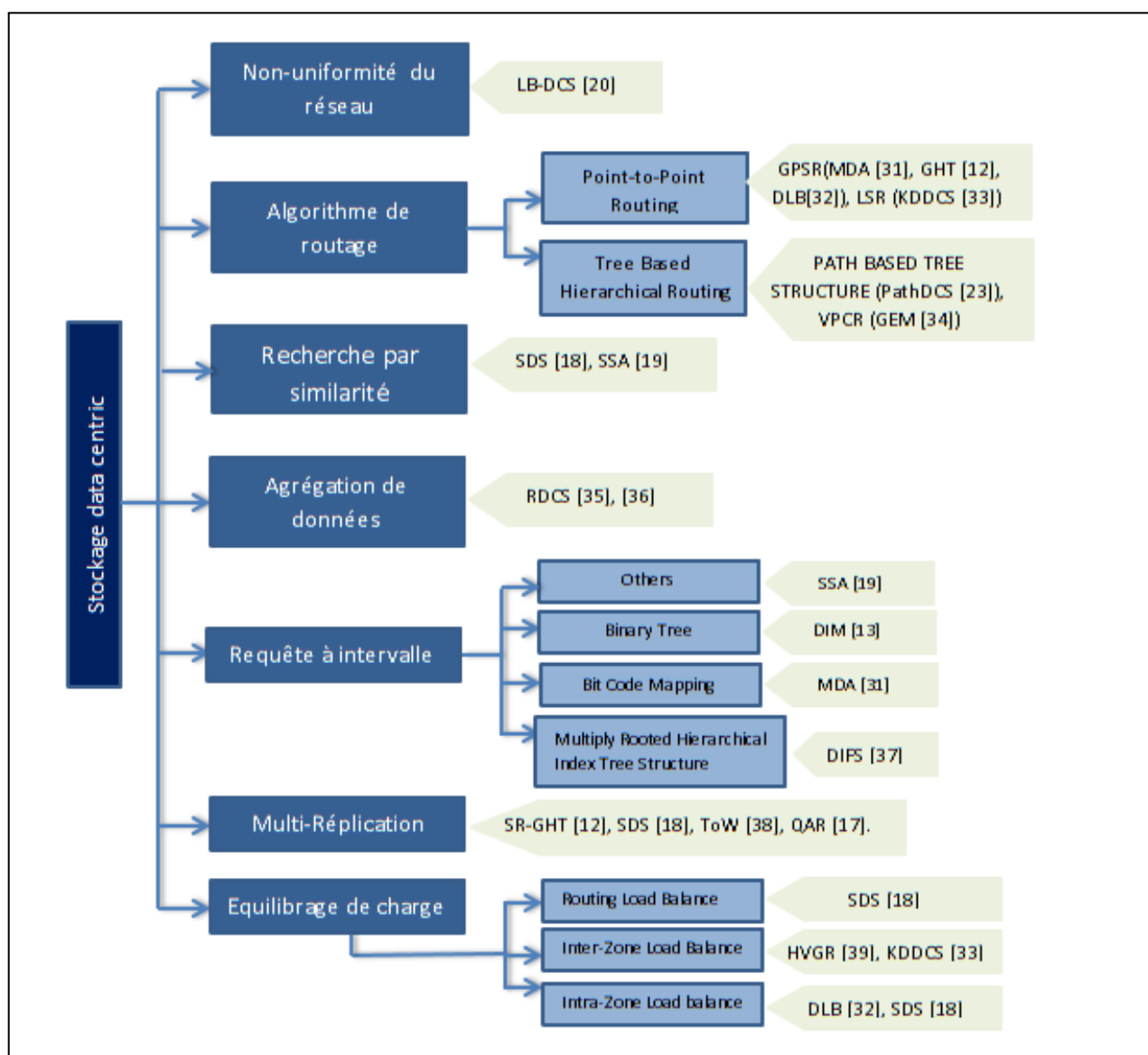
D'autres travaux ont été proposés dans le domaine DCS afin d'améliorer la première proposition. Voici une liste non exhaustive des travaux réalisés dans ce domaine : [17, 18, 19, 20, 21, 22, 23, 24].

Des états de l'art sont faits sur les travaux réalisés dans le domaine stockage data-centric dans [15, 17].

Dans [17], les auteurs ont décrit brièvement la contribution des travaux antérieurs dans le domaine DCS. Ils ont classifié ces travaux en trois catégories : multi-réplication, politique de stockage et protocole de routage.

Dans [15], Khandakar et al ont classé les travaux existants selon les différents challenges que ces travaux tentent de résoudre. Ces challenges sont : la requête à intervalle, l'agrégation de données, l'équilibrage de charge, la recherche par similarité, la réplication de données, algorithme de routage et la non-uniformité du réseau. Selon les auteurs, un travail portant sur tous ces défis ne semble pas avoir encore eu lieu. La figure 2 résume cette classification.

Un défi très important dans ce domaine n'est pas mentionné dans [15]. Il s'agit de la stratégie de choix de nœuds de stockage.



Classification des DCS.

La figure 2 présente une classification des schémas de stockage data centric faite par Khandakar et al.

B. Structures de données distribuées et scalables

Les Structures de Données Distribuées et Scalables (SDDS) sont une nouvelle classe de structures introduites vers 1993 par le Pr Litwin au CERIA (Centre de Recherche en Informatique Appliquée) [5, 6] spécifiquement pour la gestion de fichiers en mémoire centrale distribuée d'un multiordinateur. Elles fournissent un mécanisme général d'accès à des données réparties dynamiquement.

Les fichiers SDDS sont structurés en enregistrements identifiés par des clés. Ces fichiers peuvent s'étendre dynamiquement, au fur et à mesure des insertions, d'un seul site de stockage à n'importe quel nombre de sites. Des stations appelées clients peuvent accéder aux données stockées sur les serveurs. Les Clients gardent des paramètres pour le calcul des

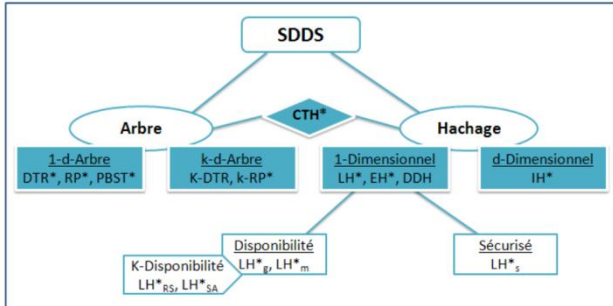
Le placement de données et son évolution sont transparents pour les applications. Celles-ci appellent en effet les clients SDDS qui gèrent l'accès aux serveurs comme s'il s'agissait de structures de données classiques. Les SDDS supportent néanmoins le traitement parallèle et assurent potentiellement par leur conception des temps d'accès aux données beaucoup plus courts que ceux aux fichiers traditionnels. Les SDDS offrent aussi une capacité de stockage potentiellement illimitée, et assurent des temps d'accès beaucoup plus courts que les structures traditionnelles. Grâce à ces caractéristiques, les SDDS offrent des performances de traitement supérieures à celles des structures de données traditionnelles. Les SDDS sont caractérisés par la scalabilité, la distribution et la disponibilité.

Compte tenu de leur philosophie, les SDDS constituent une nouvelle génération de structures de données régie par les axiomes suivantes :

- L'absence de répertoire central d'accès.
- L'extension de fichier doit être incrémentale et transparente à l'application.
- Chaque client supporte le logiciel propre à la SDDS et gère notamment sa propre image du fichier.

- Chaque serveur est capable de détecter une erreur d'adressage le concernant.
- En outre, les SDDS supportent le traitement parallèle.

Selon la stratégie de répartition des données on peut distinguer deux grandes familles de SDDS. Les SDDS basées sur la distribution par les arbres (RP* [25], DRT* [26], PBST* [4], etc.) et les SDDS basées sur la distribution par hachage (DDH [27], EH* [28], IH* [29], etc.). Cette classification est décrite dans la figure 3.



Classification des SDDS.

Dans cette session, nous avons montré l'intérêt que peut apporter les SDDS au monde actuel de l'informatique où les architectures distribuées sont le chemin qui s'impose.

Dans la session suivante nous allons présenter PBST*.

III. PBST* (PARTITIONED BINARY SEARCH TREE)

A. Description de PBST*

PBST* [4, 30] est une structure de données dédiée aux environnements distribués. Elle consiste à distribuer le fichier conformément au principe des SDDS. Elle se base sur le modèle client/serveur.

Comme toutes les SDDS, PBST* est distribué sur plusieurs serveurs. Chaque serveur S contient un ensemble d'enregistrements "case" organisés sous forme d'un arbre de recherche binaire équilibré et un intervalle [a, b].

Dans PBST* il existe deux types de serveur

- Serveur de données : il contient un arbre de données et l'adresse de son serveur père.
- Serveur de données index : il contient un arbre de données et l'adresse vers son serveur père ainsi que toutes les adresses de ses serveurs fils.

Le client PBST* a une image partielle ou complète. Cette image est un arbre de recherche binaire où chaque nœud contient l'adresse et l'intervalle de clés des serveurs qui sont déjà visités par ce client. Au départ cette image contient un seul nœud qui représente le serveur racine avec sa plage de clés qui est $]-\infty, +\infty[$. Cette image change à chaque erreur d'adressage.

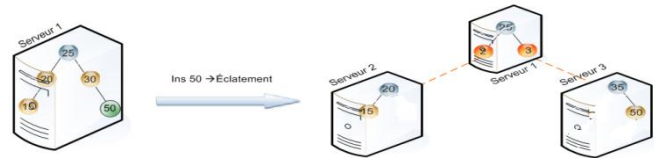
Le modèle PBST* est défini par le paramètre de partitionnement (n): un serveur PBST* contient au maximum (n-1) enregistrements.

Ce paramètre agit sur le taux de chargement des serveurs et permet la réorganisation du fichier.

B. Evolution du fichier PBST*

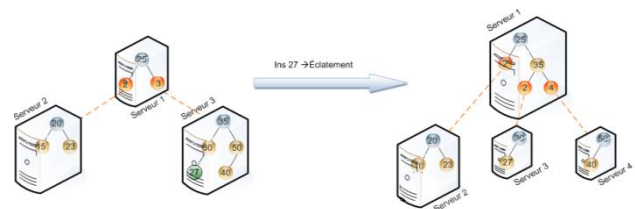
Initialement le système contient un seul serveur "serveur1" vide avec l'intervalle $]-\infty, +\infty[$, qui représente le serveur racine du fichier. Suite aux opérations d'insertions et de suppressions sur le fichier, des éclatements et des fusions des serveurs peuvent avoir lieu.

Au début, toutes les requêtes se font au niveau du serveur1 jusqu'à sa saturation. Deux nouveaux serveurs sont alors alloués, le serveur1 garde toujours le nœud racine de l'arbre. L'un des deux nouveaux serveurs reçoit le sous arbre droit et l'autre reçoit le sous arbre gauche. C'est le cas de la figure 4.



Eclatement d'un serveur racine.

Si un serveur non racine atteint le paramètre de partitionnement un nouveau serveur est alloué. Ce dernier reçoit la moitié de l'arbre du serveur saturé et la clé médiane de l'arbre monte au niveau du père de ce serveur. Ceci peut entraîner des éclatements en cascade.



Eclatement d'un serveur non racine.

Dans la figure 5, après l'insertion de la clé 27, le serveur3 atteint le paramètre de partitionnement $n=5$, donc un nouveau serveur est alloué qui est le serveur4, il reçoit la moitié de l'arbre du serveur3 et la clé médiane (dans ce cas la clé médiane=35) de ce dernier monte au niveau du serveur1 qui est le serveur père.

Le client dispose d'une image lui permettant d'accéder à un sous ensemble de serveurs. De nouveaux serveurs sont rajoutés à cette image par le biais des messages correctifs envoyés par ces serveurs. Grace à cette image, le client peut accéder directement à la partition sur laquelle il désire réaliser des opérations sans passer par le serveur racine. La figure 6 montre un exemple d'image client.

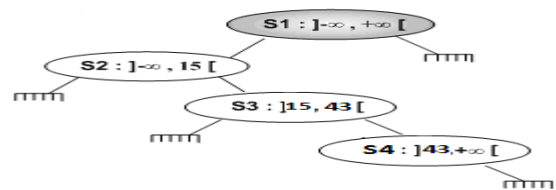


Image client.

IV. MODELE PROPOSE

Notre contribution consiste à proposer un nouveau protocole de stockage de données basé sur PBST*, une

variante des structures de données distribuées et scalables (SDDS), pour optimiser le stockage et la collection des données par la station de base (puits).

Nous nous intéressons aux modèles de transmission suite à des requêtes. Cette catégorie peut inclure les modèles de détection d'événements : dans ce cas, la requête sera déferée dans le temps, c'est-à-dire, on envoie la requête aux capteurs avec une durée de vie et on répond lorsqu'on détecte un événement (exemple : le protocole de routage : directed diffusion [10]).

Le modèle de transmission suite à des requêtes est très intéressant car dans certaines applications les mesures envoyées par les capteurs ne sont pas toutes utiles pour la station de base. En effet, dans certaines applications, un utilisateur ne veut pas visualiser toutes les mesures de tous les capteurs dans un réseau. Par exemple, il peut poser une requête pour voir à quel moment, la température sur une zone a dépassé 60°C. Dans ce cas, toutes les transmissions de température qui sont inférieures à 60°C sont inutiles. Il vaut mieux stocker ces mesures sur les capteurs que de les transmettre à la station de base.

Dans notre modèle, on suppose que le réseau est divisé en plusieurs zones. Chaque zone est responsable d'un type de données.

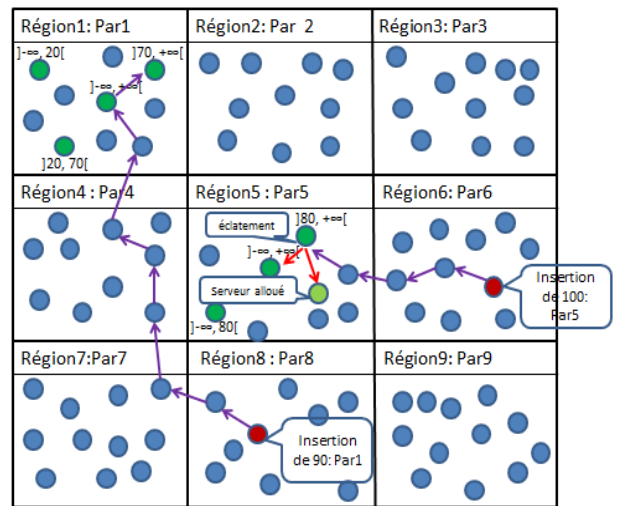
Le capteur qui détecte un événement est un client, il envoie cet événement vers un autre capteur (serveur) pour le stockage de ce dernier. La station de base (puits) est un client collecteur. Elle fait de requêtes de recherche (récupération). Un client (capteur) fait des requêtes d'insertions des événements détectés.

A. Protocole d'insertion

Lorsqu'un capteur détecte un événement, il l'envoie à la zone qui est responsable de ce type d'événements. Le premier capteur de cette zone qui reçoit la requête, redirige cette dernière vers le serveur adéquat, ce serveur peut aussi rediriger la requête en cas d'erreur d'adressage. Le serveur dont l'intervalle contient l'événement détecté, l'insère dans son arbre.

L'insertion peut entraîner un éclatement ou des éclatements en cascade. Lors d'un éclatement, le nouveau serveur alloué est choisi selon son niveau d'énergie ainsi que la distance qui le sépare du serveur éclaté. On met aussi à jour les capteurs voisins de serveur éclaté en changeant l'intervalle du serveur éclaté et en ajoutant les informations concernant le nouveau serveur alloué.

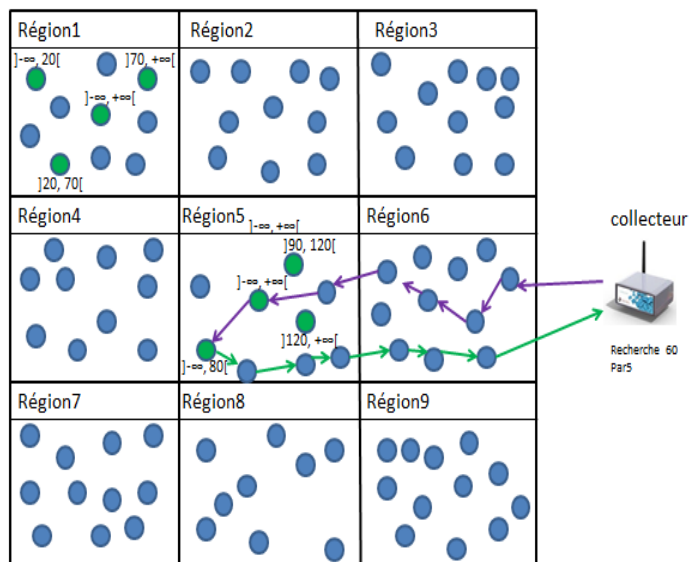
Dans la figure 7, deux nœuds producteurs captent deux nouveaux événements de différents types. Chaque nœud envoie sa donnée à la région responsable de l'événement détecté. Le capteur qui a détecté la donnée (90) de type Par1, l'envoie à la région1. A l'intérieur de cette région, cette événement est redirigé jusqu'à atteindre le serveur dont l'intervalle contient 90. Le même traitement est fait sur l'événement de clé 100 de type Par5. L'insertion de cet événement a entraîné un éclatement et donc l'allocation d'un nouveau serveur.



Protocole d'insertion.

B. Protocole de recherche

Les requêtes de recherche sont toujours envoyées par le collecteur. Lorsque le collecteur désire faire une recherche sur une donnée de type spécifique, il envoie une requête de recherche à la région responsable de ce type. Cette requête peut être redirigée à l'intérieur de la zone jusqu'à atteindre le serveur adéquat. Lorsque la requête arrive au niveau du serveur dont l'intervalle contient la donnée recherchée, il répond favorablement si la donnée existe ou défavorablement dans l'autre cas. Le chemin de la réponse peut être différent de celui de la requête.



Protocole de recherche.

La figure 8 montre le déroulement d'une requête de recherche. Le collecteur envoie la requête de recherche vers la région 5. Cette est redirigé vers le serveur adéquat. Ce dernier répond au collecteur en prenant probablement un autre chemin.

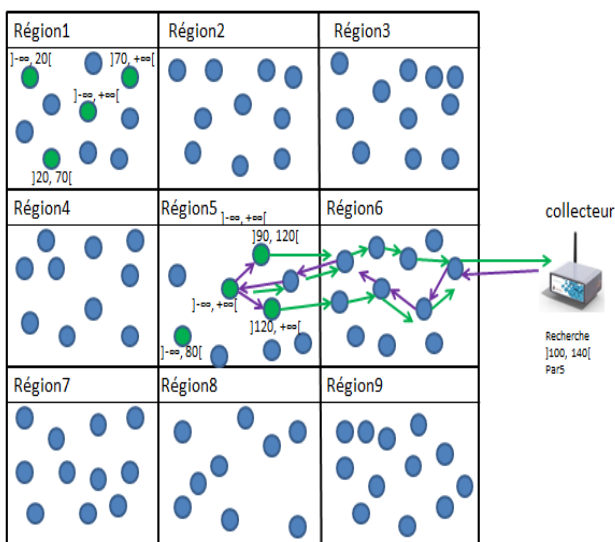
C. Protocole de la requête à intervalle

La requête à intervalle est un autre défi pour les systèmes de stockage distribué de données. En effet, un utilisateur peut s'intéresser à un intervalle plutôt qu'à une seule valeur spécifique. Par exemple, la pollution de l'air

peut se produire si le niveau de monoxyde de carbone est dans l'intervalle [30 L / mol, 90 L / mol]. Une requête possible est de trouver tous les points où le niveau de monoxyde de carbone appartient à cet intervalle. Avec ce type de requête, les utilisateurs peuvent améliorer leur efficacité de la recherche. L'exemple précédent illustre une utilisation très importante des requêtes à intervalle.

Notre proposition permet de faire des requêtes à intervalle car les données sont réparties de manière ordonnée sur l'ensemble de nœuds de stockage.

Comme l'opération de recherche, la requête à intervalle est toujours envoyée par le collecteur. Elle est spécifique pour un type particulier de données. Le collecteur envoie la requête à intervalle à la région dont le type est le même que celui de la requête. La requête peut être redirigée par les capteurs de la région vers le serveur susceptible d'inclure la totalité de l'intervalle.



Protocole de requête à intervalle.

Dans la figure 9, le collecteur envoie une requête à intervalle à la région5 car c'est elle qui est responsable du Par5. Lorsque cette requête arrive au niveau du serveur dont l'intervalle couvre la plage de la requête celui-ci la redirige vers les serveurs adéquats. Il peut aussi répondre à la requête s'il contient des valeurs appartenant à l'intervalle de cette dernière.

V. CONCLUSION

Le stockage distribué de données dans les WSN est une thématique de recherche en pleine expansion. Dans cet article, nous avons proposé une nouvelle méthode de stockage de données distribué pour les WSN. Cette méthode est basée sur PBST* qui permet de distribuer les données sur l'ensemble de ressources de stockage de manière ordonnée et équilibrée.

Cette solution peut traiter les challenges suivants:

- Requêtes à intervalle : PBST* est un arbre de recherche binaire partitionné, cette méthode stocke les données de manière ordonnée. Par conséquent, les requêtes à intervalles peuvent être exécutées rapidement.

- l'équilibrage de charge : les tests de simulation montre que PBST* distribue les données de manière équilibré sur l'ensemble des ressources de stockage (le taux de chargement des serveurs avoisine de 70%).

REFERENCES

- [1] S. Tilak, N.B. Abu-Ghazaleh et W.B. Heinzelman, "A taxonomy of wireless microsensor network models.", ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 6, no. 2, pp. 28-36, April 2002.
- [2] C.F. Garcia-hernandez, P.H. Ibarguengoytia-gonzalez, J. Garcia-hernandez, J.A. Pérez-diaz, "Wireless Sensor Networks and Applications: a Survey", International Journal of Computer Science and Network Security, March 2007
- [3] A. Boukerche, R. W. N. Pazzi et R. B. Araujo, "A fast and reliable protocol for wireless sensor networks in critical conditions monitoring applications.", Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems (MSWiM '04), pp. 157-164, 2004.
- [4] A. Chikhaoui, D.E Zegour, W.K Hidouci. "PBST*: une nouvelle variante pour les SDDS." Rencontre sur la recherche (R2I), Juin 2011, Tiziouzou, Algérie.
- [5] W. Litwin, M.A. Neimat, D. Schneider: "LH*: Linear Hashing for Distributed Files." ACM-SIGMOD Int. Conf. On Management of Data, 1993.
- [6] W. Litwin, M.A. Neimat, D. Schneider: "LH*: A Scalable Distributed Data Structure", Submitted for journal publ, 1993.
- [7] S. Shenker, S. Ratnasamy, R.G.B. Karp et D. Estrin, "Data-centric storage in sensornets.", In the 1st ACM SIGCOMM Workshop on Hot Topics in Networks (HotNets'02), October 2002.
- [8] G.J. Pottie, W.J. Kaiser, "Wireless integrated network sensors." Commun. ACM 2000, pp. 51-58, 2000.
- [9] R. Szweczyk, J. Polastre, A. Mainwaring, D. Culler, "Lessons from a Sensor Network Expedition.", In Proceedings of European Workshop Wireless Sensor Network, , pp. 307-322, January 2004, Berlin, Germany.
- [10] C. Intanagonwivat, R.Govindan, D.Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks." In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 56-67, August 2000, Boston, USA.
- [11] F. Ye, G. Zhong, S. Lu, L. Zhang, "Gradient broadcast: A robust data delivery protocol for large scale sensor networks." Wirel. Netw.2005, pp. 285-298, 2005.
- [12] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R.Govindan, S. Shenker, "GHT: A Geographic Hash Table for Data-centric Storage.", In Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, pp. 78-87, 28 September 2002, Atlanta, GA, USA.
- [13] X. Li, Y.J. Kim, R.Govindan, W. Hong, "Multi-dimensional Range Queries in Sensor Networks.", In Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, pp. 63-75, November 2003, Los Angeles, USA
- [14] D. Ganesan, A. Cerpa, W. Ye, Y. Yu, J. Zhao, D. Estrin, "Networking issues in wireless sensor networks.", J. Parallel Distrib. Comput.2004, pp. 799-814, 2004.
- [15] A. Khandakar, A.G. Mark, "Techniques and Challenges of Data Centric Storage Scheme in Wireless Sensor Network", Journal of Sensor and Actuator Networks, June 2012.
- [16] B. Karp, H.T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks." MobiCom 2000, pp. 243-254, Boston, 2000.
- [17] A.C. Rumin, M.U. Pascual, R.R. Ortega, D.L. López, "Data centric storage Technologies: Analysis and enhancement." Sensors 2010, pp. 3023-3056, 2010.
- [18] H. Shen, L. Zhao, Z. Li, "A distributed spatial-temporal similarity data storage scheme in wireless sensor networks." IEEE Trans. Mob. Comput, pp. 982-996, 2011.

- [19] Y. Chung, I.F. Su, C. Lee, "An efficient mechanism for processing similarity search queries in sensor networks." *Inf. Sci.*, pp. 284–307, 2011.
- [20] M. Albano, S. Chessa, F. Nidito, S. Pelagatti, "Dealing with nonuniformity in data centric storage for wireless sensor networks." *IEEE Trans. Parallel Distrib. Syst.*, pp. 1398–1406, 2011.
- [21] K. Seada, A. Helmy, "Rendezvous Regions: A Scalable Architecture for Service Location and Data-Centric Storage in Large-Scale Wireless Networks." *Parallel Distrib. Process. Symposium, Int.* 2004.
- [22] Y. Zhao, Y. Chen, S. Ratnasamy, "Load Balanced and Efficient Hierarchical Data-Centric Storage in Sensor Networks." In *5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2008. SECON '08, pp. 560–568, 2008, New York, USA.
- [23] C.T. Ee, S. Ratnasamy, S. Shenker, "Practical data-centric storage." In *NSDI'06: Proceedings of the 3rd conference on Networked Systems Design & Implementation*, USENIX Association: Berkeley, USA, 2006.
- [24] N.L. Thang, Y. Wei, B. Xiaole, X. Dong, "A dynamic geographic hash table for data-centric storage in sensor networks." In *IEEE Wireless Communications and Networking Conference, WCNC 2006*, pp. 2168 – 2174, New York, USA, 2006.
- [25] W. Litwin, M.A. Schneider, "RP*: A family of Order-Preserved Scalable Distributed Data Structure." *20th International Conference, On Very Large Data Bases (VLDB)*, 1994.
- [26] B. Kroll, P. Widmayer : "Distributing a Search Tree Among a Growing." In *ACM-SIGMOD International Conference On Management of Data*, pp. 265-276, 1994.
- [27] D. Devin, "Design and implementation of DDH: A distributed dynamic hashing algorithm." In *Processing of the 4th Foundation of Data Organization and Algorithms (FODO)*, 1993.
- [28] V. Hilford, F.B. Bastani and B. Cukic, "EH*: Extendible Hashing distributed." 1997.
- [29] D.E. Zegour and D. Boukhelef : "IH*: Hachage Linéaire Multidimensionnel Distribué et Scalable." *Conférence Africaine de Recherche en Informatique, CARI 2002*, Octobre 2002, Yaoundé, Cameroun.
- [30] A. Chikhaoui, D.E. Zegour, W.K. Hidouci. Towards dynamic data placement in grid. *International Conference on Information Systems and Technologies (ICIST)*, Avril 2011, Tebessa, Algérie
- [31] W.H. Liao, C.C. Chen." Data storage and range query mechanism for multi-dimensional attributes in wireless sensor networks." *Communications* 2010.
- [32] W.H. Liao, K.P. Shih, W.C. Wu. "A grid-based dynamic load balancing approach for data-centric storage in wireless sensor networks." *Comput. Electr. Eng.*, pp. 19-30 2010.
- [33] M. Aly, K. Pruhs, P.K. Chrysanthis. "KDDCS: A Load-Balanced In-Network Data-Centric Storage Scheme for Sensor Networks." In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, Arlington, VA, USA*, pp. 317–326, November 2006.
- [34] J. Newsome, D. Song. "GEM: Graph Embedding for Routing and Data-Centric Storage in Sensor Networks without Geographic Information." In *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, Los Angeles, CA, USA*, pp. 76–88, November 2003.
- [35] A. Ghose, J. Grossklags, J. Chuang. "Resilient Data-Centric Storage in Wireless Ad-Hoc Sensor Networks." In *Proceedings of the 4th International Conference on Mobile Data Management, Melbourne, Australia*, pp. 45-62, January 2003.
- [36] A. Cuevas, M. Uruena, R. Cuevas, R. Romeral. "Modelling data-aggregation in multi-replication data centric storage systems for wireless sensor and actor networks." *Communications* 2011.
- [37] B. Greenstein, D. Estrin, R. Govindan, S. Ratnasamy, S. Shenker. "DIFS: A Distributed Index for Features in Sensor Networks." In *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, Anchorage, AK, USA*, pp. 163-173, May 2003.
- [38] Y.J. Joung, S.H. Huang. "Tug-of-War: An Adaptive and Cost-Optimal Data Storage and Query Mechanism in Wireless Sensor Networks." *Lect. Note. Comput. Sci.* 2008.
- [39] Z. Yao, C. Yan, S. Ratnasamy. "Load Balanced and Efficient Hierarchical Data-Centric Storage in Sensor Networks." In *Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, San Francisco, CA, USA*, pp. 560–568, June 2008.

Bi-Clustering Algorithm Using Formal Concept Analysis

Amina Houari

University of Tunis El Manar
Faculty of Sciences of Tunis
LIPAH
aminahouari.fst@gmail.com

Wassim Ayadi

University of Tunis
Higher School of Sciences
and Technologies of Tunis
LATICE
University of Angres, LERIA
wassim.ayadi@gmail.com

Sadok Ben Yahia

University of Tunis El Manar
Faculty of Sciences of Tunis
LIPAH
sadok.benyahia@fst.rnu.tn

Abstract— Identifying groups of genes that show similar behavior under a subset of conditions is known to be hard due to the huge volume of gene expression data. Thus, there is a compelling in a great need for novel techniques to perform such task.

In this paper, we introduce a new algorithm for extracting biclusters from microarray data. Our algorithm relies on Formal Concept Analysis, which has been shown to be an efficient methodology for biclustering binary data. The performance of our algorithm is evaluated on real DNA microarray dataset.

I. INTRODUCTION

One of the main challenges in bioinformatic is the design of efficient algorithms capable of clustering biological data. Such clusters are of valuable importance for researchers on cancer studies like cancer occurrences, specific tumor subtypes, or cancer survivals. This task is achieved through a particular type of clustering known as "Biclustering". In this respect, the biclustering of numerical data has become an important task of biological analysis of gene expression data.

Cheng and Church [1] were the first to apply biclustering to gene expression data. The main objective is to identify maximal sub-matrices of genes and conditions such that the genes express highly correlated behaviors over a range of conditions. Biclustering is known to be NP-hard [1]. Clustering genes according to their expression profiles is important to extract knowledge from microarray data. To do so, gene expression data is arranged in a data matrix where rows represent genes and columns represent conditions.

Several surveys of biclustering algorithms have been given. In [2], the authors distinguish two main classes of biclustering algorithms: Systematic search algorithms and stochastic search algorithms (meta-heuristic algorithms). The survey given by Madeira and Oliveira [3] is also worth of mention in the latter, biclusters are defined to be a subset of genes and conditions such that the subset of genes exhibits similar fluctuations under a given subset of conditions. In addition, biclusters can overlap on genes and on conditions.

The above mentioned biclustering algorithms exhibit some limitations due to their heuristic nature, and therefore, are not guaranteed to find the optimal set of biclusters. While Biclustering algorithms discussed in the literature and using formal concept analysis have the tendency to either focus on

one type of biclusters or refrain from validating them biologically (like in [4] and [5]).

In this paper, we introduce a new algorithm for biclustering of DNA microarray data, called BiFCA. Our algorithm operates in three steps. First, BiFCA discretize the original microarray data in order to observe the profile of each gene through all pairs of conditions. Second, BiFCA is based on Formal Concept Analysis [6] which is a mathematical framework for allowing to derive implicit relationships from a set of objects and their attributes. Third, to avoid the high overlap between biclusters, we use the bond measure.

The remainder of the paper is organized as follows: Section 2 offers a state of the art study of biclustering. Section 3 is dedicated to the description of the BiFCA algorithm. The application of our approach on Yeast cell-cycle dataset is provided in Section 4. The conclusion and perspectives are sketched in Section 5.

II. BICLUSTERING

A. Notations

In the following, we present the main notions that will be of use in the remainder;

Definition 1 (Biclustering)

The biclustering problem focuses on the identification of the best biclusters of a given dataset.

In Biclustering of DNA microarray dataset, the latter is represented by a data matrix, where each cell represents the gene expression level of a gene under an experimental condition. Formally, a bicluster can be defined as follows:

$I = \{1, 2, \dots, n\}$ a set of indices of n genes, $J = \{1, 2, \dots, m\}$ a set of indices of m conditions, and $M(I, J)$ the data matrix associated with I and J .

Definition 2 (Biclusters)

A bicluster is a subset of genes associated with a subset of conditions in which these genes are co-expressed.

The bicluster associated with the matrix $M(I, J)$ is a couple (A, B) , such that $A \subseteq I$ and $B \subseteq J$, and (A, B) is maximal (if there does not exist a bicluster (C, D) with $A \subseteq C$ or $B \subseteq D$).

Definition 3 (Formal Concept Analysis)

Formal concept analysis (FCA) [6] is a method of data analysis with growing popularity across various domains. FCA analyzes data which describes the relationship between a particular set of objects G and a particular set of attributes M .

Wille introduced Formal Concept Analysis in 1982 [6]. A formal concept is a set of objects, *extent*, for which apply a set of attributes, *intent*. FCA provides a tool of classification and analysis, the central concept is the concept of a lattice.

TABLE I. EXAMPLE OF A FORMAL CONTEXT

	A	B	C	D	E
1	×		×	×	
2		×	×		×
3	×	×	×		×
4		×			×
5	×	×	×		×

Definition 4 (Formal context)

A formal context is a triple $K=(G, M, R)$ where G is a set of objects, M is a set of attributes and the binary relation $R \subseteq G \times M$ shows which objects have which attributes.

A formal context can be represented by a cross- table (Table I). For $A \subseteq G$, we define : $A' = \{m \in M \mid \forall g \in A, (g, m) \in R\}$

and dually for $B \subseteq M$: $B' = \{g \in G \mid \forall m \in B, (g, m) \in R\}$

Roughly speaking, A' is the set of all attributes common to the objects of A , while B' is the set of all objects that have all attributes in B .

Definition 5 (Formal concept)

A pair (A, B) is a formal concept of (G, M, R) if and only if $A \subseteq G, B \subseteq M, A' = B$ and $A = B'$

In other words, (A, B) is a formal concept if the set of all attributes shared by the objects of A is identical to that of B and vice versa.

A is then called the *extent* and B the *intent* of the formal concept (A, B) .

Definition 6 (The bond measure)

The bond measure of two concepts' intents I_1, I_2 is defined as follows [7]:

$$\text{bond}(I_1, I_2) = \frac{\text{Supp}(\wedge I_1, I_2)}{\text{Supp}(\vee I_1, I_2)}$$

So, we can redefine the bond measure as follows:

$$\text{bond}(I_1, I_2) = \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|}$$

B. Related work

The biclustering allows simultaneous clustering of both genes and conditions of the data matrix. Biclustering was introduced in [8], then used on microarray gene expression data in [1] by Cheng and Church (CC) to find biclusters with low variance as defined by the Mean Squared Residue (MSR). Since its inception, a large number of biclustering algorithms were proposed.

In the survey done in [2], the authors distinguished two main classes of biclustering algorithms: Systematic search algorithms and stochastic search ones (meta-heuristic algorithms).

The systematic search algorithms are based on:

1) *Divide-And-Conquer (DAC) based approach*: Generally, this method repeatedly splits the problem into smaller ones with similar structures to the original problem, until these subproblems become smaller enough to be solved directly. The solutions to the subproblems are then combined to create a solution to the original problem [2]. In [9], the authors partitioned the discretized data matrix into three submatrices, one of which contains only 0-cells. The algorithm is then recursively applied to the remaining two submatrices, and comes to an end if the current matrix represents a bicluster which contains only 1s. Algorithms adopting this approach are given in [9] and [10].

2) *Greedy Iterative Search (GIS) based approach*: They construct a solution in a step-by-step way using a given quality criterion. Decisions made at each step are based on information at hand without worrying about the impact these decisions would have in the future. Moreover, once a decision is made, it becomes irreversible and is never reconsidered [2]. For instance, in [11], the authors proposed the Time Series Biclustering algorithm, which refers to CC algorithm by restricting it to add and remove columns. Algorithms adopting this approach are given in [12].

3) *Biclusters Enumeration (BE) based approach*: The enumeration algorithm enumerates all the solutions for the original problem. The enumeration process is generally represented by a search tree [2]. In [13], they use a Bicluster Enumeration Tree (BET) to find all the biclusters (nodes), reachable from the root of the BET, by adopting an evaluation function. Whereas, in [14], the authors proposed an algorithm named Iterative Signature Algorithm (ISA), in which they defined a bicluster as a set of co-regulated genes which associates a set of co-regulated conditions. Algorithms adopting this approach are given in [15] and [16].

While Stochastic search algorithms are based on:

1) *The Neighborhood Search (NS) based approach*: It starts with an initial solution and then moves iteratively to a neighboring solution thanks to the neighborhood

exploitation strategy[2]. Algorithms adopting this approach are given in [17] and [18].

2) *Evolutionary Computation(EC) based approach*: it is based on the natural evolutionary process such as population, reproduction, mutation, recombination, and selection [2]. In [1], the authors tried to find biclusters with low variance as defined by the Mean Squared Residue(MSR). Algorithms adopting this approach are given in [19].

3) *Hybrid (H) based approach*: it tries to combine the neighborhood search and the evolutionary approaches. This hybrid approach is known to be quite successful in solving many hard combinatorial search problems[2]. Algorithms adopting this approach are given in [20] and [21].

Even through their number is high, the above mentioned biclustering algorithms exhibit some limitations due to their heuristic nature, and therefore, are not guaranteed to find the optimal set of biclusters.

To palliate such a drawback, the approaches based on the Formal Concept Analysis were introduced. In [22], the authors proposed a new approach, called FIST, for extracting bases of extended association rules and conceptual biclusters, using the frequent closed itemsets. While in [23], they use a single threshold, where expression values greater than this threshold are represented by 1, otherwise by 0.

Whereas, in [5], they used the scaling of numerical data and consider that formal concepts are the groups of genes whose expression values are in the same intervals for a sub-set of conditions.

In[4], the authors refer to the algorithm done in [5], using Triadic Concept Analysis [24] in order to extract biclusters with similar values.

These biclustering algorithms using formal concept analysis have the tendency to either focus on one type of biclusters or refrain from biological validation.

III. ALGORITHM DESCRIPTION

The BiFCA biclustering algorithm is based on formal concept analysis in order to extract formal concepts. BiFCA operates in four main steps, we start with discretizing the initial numerical data matrix into a -101 data matrix¹ which represents the relation between all conditions for the gene set in the gene expression matrix, then we discretize the -101 data matrix into a binary one in order to extract formal concepts.

Finally, we compute the bond measure which is defined as the ratio between a conjunctive support of a concept and its disjunctive support, and we consider only those having the bond measure not exceeding a given threshold α , done, in order to remove concepts that have high overlapping.

The pseudo-code description of algorithm BiFCA is shown in Algorithm 1.

Algorithm 1 BiFCA Algorithm

```

1: Input: M, minbond; Output:  $\beta$ 
2: Discretize M using Equation (1) to obtain M2.
3: Discretize M2 using Equation (2) to obtain M3.
4: Extract all formal concepts.
5: for each two biclusters  $B_1, B_2$  do
6: if  $Bond(B_1, B_2) < minbond$  then  $\beta = \beta \cup \{B_1, B_2\}$ 
7: else  $\beta = \beta \cup \{B_1 \text{ or } B_2\}$  // which have a higher number of
conditions
8: endfor
9: Return  $\beta$ 

```

EXAMPLE OF GENE EXPRESSION MATRIX

	c1	c2	c3	c4	c5	c6
g1	10	20	5	15	0	18
g2	20	30	15	25	26	25
g3	23	12	8	15	20	50
g4	30	40	25	35	35	15
g5	13	13	18	25	30	55
g6	20	20	15	8	12	23

A. From numerical data to -101 data matrix

Our method first applies a preprocessing step to transform the original data matrix M into a -101 data matrix M2 (Ligne 2). This step aims to highlight the trajectory patterns of genes. According to both [25] and [26], in microarray data analysis, we add genes into a bicluster (cluster) whenever their trajectory patterns of expression levels are similar across a set of conditions.

Interestingly enough, our proposed discretization step keeps track of the profile shape over conditions and preserves the similarity information of trajectory patterns of the expression levels.

Before applying the formal concepts extraction algorithm, we must first discretize the initial data matrix. The discretization process outputs the -101 data matrix. It consists in combining in pairs, for each gene, all the conditions between them. Indeed the -101 data matrix gives an idea about the profile. Furthermore, one can have a global view of the profile of all conditions between them.

In our case, each column of the -101 data matrix represents the meaning of the variation of genes between a pair of conditions of M. It offers useful informations for the identification of biclusters.

Formally the matrix M2(-101 data matrix) is defined as follows :

$$M2 = \begin{cases} 1 & \text{if } x_1 < x_2 \\ -1 & \text{if } x_1 > x_2 \\ 0 & \text{if } x_1 = x_2 \end{cases}$$

With: $x_1 = M[i, l]$; $x_2 = M[i, l2]$;

and $i \in [1 \dots n]$; $l \in [1 \dots m-1]$; $l2 \in [i+1 \dots m]$;

Let us consider the dataset given by table II. Using equation (1) we represent the -101 data matrix (see table III).

¹ The -101 data matrix is the matrix obtained after the discretization of the original data matrix.

THE -101 DATA MATRIX (M2)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
g1	1	-1	1	-1	1	-1	-1	-1	-1	1	-1	1	-1	1	1
g2	1	-1	1	1	1	-1	-1	-1	-1	1	1	1	1	0	-1
g3	-1	-1	-1	-1	1	-1	1	1	1	1	1	1	1	1	1
g4	1	-1	1	1	-1	-1	-1	-1	-1	1	1	-1	0	-1	-1
g5	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
g6	0	-1	-1	-1	1	-1	-1	-1	1	-1	-1	1	1	1	1

B. From -101 data to binary data matrix

Let *M2* be a -101 data matrix (e.g Table III), in order to build the binary data matrix(Ligne 3), we compute the average number of repetitions for each column in the matrix *M2*, for example for the column *C1* we have: maxrepeat² is 3 and corresponds to the maxvalue 1. minrepeat³ is 1 and corresponds to the minvalue -1. And mediamrepeat = 2 and corresponds to the value 0. So, the *average value* is 0, passing to the binary matrix the column *C1* becomes as sketched in table IV.

EXAMPLE OF A COLUMN DISCRETIZATION

	C1
g1	0
g2	0
g3	0
g4	0
g5	1
g6	1

It is better to choose the mean value since the maximum will produce a huge number of overlapping biclusters, while the minimum value generates biologically none-valid biclusters.

Then, we define the binary matrix as follows:

$$M3 = \begin{cases} 1 & \text{if } x_i = \text{averagevalue} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Using Equation(2) we obtain the binary matrix (see Table V).

THE BINARY DATA MATRIX (M3)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
g1	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0
g2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
g3	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0
g4	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1

g5	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
g6	1	0	1	1	0	0	0	0	0	1	1	0	0	0	0

C. Extracting Formal concepts

After preparing the binary data matrix, we move to extract formal concepts (Ligne 4) from the matrix M3(Table V).

To perform this task we divide the problem in two sub-problems:

1. Find all frequent closed itemsets.
2. Extract all Formal concepts.

Many algorithms have been proposed in the literature for finding frequent closed itemsets (for more information see [27]).In this paper, we use an efficient algorithm in order to enumerate all frequent closed itemsets (conditions) which represent *intents* of the formal concepts, then for each closed itemset it extracts the proper *extents* (genes).

By using the previous example we obtain as a result the concepts in TableVI.

THE FORMAL CONCEPTS EXTRACTED FROM THE BINARY CONTEXT

	intents (conditions)	extents (genes)
concept1	C4	g1, g3, g6
concept2	C9	g1, g2, g4
concept3	C1	g5, g6
concept4	C3, C4	g3, g6
concept5	C7, C8	g3, g5
concept6	C3, C4, C7, C8	g3
concept7	C4, C11	g1, g6
concept8	C9, C15	g2, g4
concept9	C1, C2, C6, C7, C8	g5
concept10	C1, C3, C4, C10, C11	g6
concept11	C4, C9, C11, C13	g1
concept12	C5, C9, C12, C14, C15	g4

D. The bond measure

Before computing the bond measure between two concepts' intents, we filter the concepts by taking into consideration with the number of conditions is higher than a given threshold, due to the overwhelming number of generated formal concepts.

The BiFCA algorithm is already able to identify overlapping biclusters. In order to compute the similarity between two biclusters (concepts) C1 and C2, C1= (I₁, E₁), C2= (I₂, E₂) where I_i, i = 1, 2, represents the intent and E represents the extent, we use the measure bond [7]. The latter measures the overlapping between two concepts or Biclusters.

²Maximum number of occurrences by column.
³Minimum number of occurrences by column.

In fact, for the filtering process (Ligne 5, 6, 7) we consider only biclusters with a low overlap (if biclusters have a high overlapping, they done the same biological signification) . The correlation measure bond achieves its minimum of 0 when the biclusters do not overlap at all and its maximum of 1 when they are identical.

Using the example of concept5 and concept6 (example from Table VI),we compute the bond measure :

$$\text{bond}(I_5, I_6) = \frac{|\{C_7, C_8\} \cap \{C_3, C_4, C_7, C_8\}|}{|\{C_7, C_8\} \cup \{C_3, C_4, C_7, C_8\}|}$$

$$\text{bond}(I_5, I_6) = \frac{2}{4} = 0.50$$

Taking 0.6 as threshold of bond ,we consider the both concepts C_5 and C_6 as non overlapping biclusters. While taking 0.3 as threshold of bond ,we consider only C_6 as a bicluster i.e which have the higher number of elements.

IV. EXPERIMENTAL RESULTS

In this section, we assess the BiFCA algorithm on real DNA microarray data.

All experiments were carried out on a PC equipped with a 2.53 GHz Intel(R) Core(TM)i3 CPU and 4GB of main memory, running the Linux Ubuntu 12.04.

The yeast cell-cycle is a very popular dataset in the gene expression data. In fact, it is the one of the well known organisms and the functions of each gene are well known.

We used the *Yeast cell-cycle dataset* which was described in [28], processed in [1] and is publicly available from [29], it contains 2884 genes and 17 conditions.

A. Discussion of obtained results

To evaluate the quality of the extracted biclusters and identify their biological annotations, we use *GO Term Finder*⁴ which is designed to search for the significant shared Gene Ontology (GO) terms of a group of genes. The ontologies are in the form of direct acyclic graphs where GO terms represent nodes and the relationships between them represent edges. These ontologies are organized in three domains: *biological process*, *molecular function* and *cellular component*. We show in Table VII the result of a selected set of genes for biological process, molecular function and cellular component, we report the most significant GO terms. With the first bicluster (Table VII) the genes (YBL027W, YBL072C, YBL087C, YBL092W, YBR031W, YBR048W, YBR079C, YBR084C-A, YBR181C, YBR191W, YCR031C, YDL061C, YDL075W, YDL081C, YDL082W, YDL083C, YDL130W, YDL136W, YDL191W, YDL229W, YDR012W, YDR025W, YDR064W, YDR382W, YDR418W, YDR447C, YDR450W, YDR471W, YDR500C, YER074W, YER102W, YER117W, YER131W, YGR214W, YHL001W, YHR141C, YIL069C, YJL177W, YJL189W, YJL190C, YJR094W-A, YJR123W, YJR145C, YKL056C, YKL156W, YKL180W, YKR057W, YKR094C, YLL045C, YLR048W, YLR075W, YLR167W, YLR185W, YLR325C, YLR340W, YLR344W, YLR367W, YLR388W, YLR406C, YLR441C, YLR448W, YML024W, YML026C, YML063W, YML073C, YMR121C, YMR143W, YMR146C, YMR194W, YMR230W, YMR242C, YNL067W, YNL096C, YNL162W,

YNL301C, YNL302C, YOL039W, YOL040C, YOL127W, YOR167C, YOR234C, YOR293W,

Significant GO terms for two biclusters extracted by BiFCA

	Bicluster1	Bicluster 2
Biological Process	cytoplasmic translation (p-value 1.08 e-51) ribosomal small subunit biogenesis (p-value 6.41 e-13)	ribosomal small subunit biogenesis (p-value 7.49e-13) translation (p-value 4.35e-11)
Molecular Process	structural constituent of ribosome (p-value 1.16 e-40) structural molecule kinase regulator activity (p-value 1.10 e-35)	structural constituent of ribosome (p-value 1.91e-40) kinase regulator activity (p-value 9.91e-05).
Cellular Component	ribosomal subunit (p-value 1.23 e-38) ribosome (p-value 2.13 e-36).	cytosolic ribosome (p-value 6.11e-53) cytosolic part (p-value 2.37e-40)

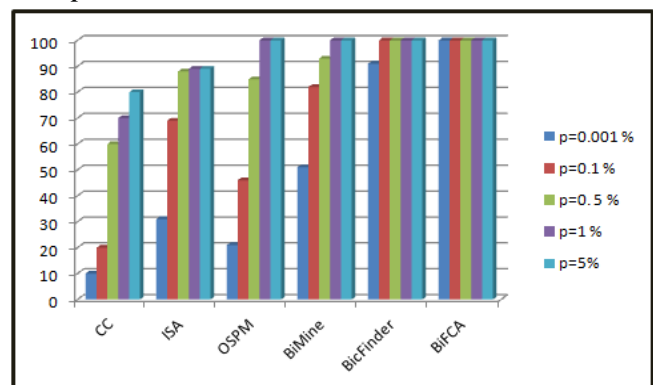
YOR312C, YOR369C, YPL081W, YPL090C, YPL143W, YPL198W, YPR043W, YPR102C, YPR163C)

concern *the cytoplasmic translation*⁵ with *p-value* of $1.08 e^{-51}$ (Highly significant) and ribosomal small subunit biogenesis.

The results on this real data set show that our proposed algorithm can identify biclusters with a high biological relevance.

On the other hand, we also use the web tool *FuncAssociate*⁶ [30] in order to compute the adjusted significance scores for each bicluster (adjusted *p-value*⁷). Then, the results of our algorithm are compared with CC[1], ISA[31], OSPM[32], BiMine[13], BicFinder [33], we report the result of the algorithms mentioned before from the values used in [34].

The obtained results for different significant scores *p* for each algorithm over the percentage of total biclusters are depicted in Figure 1. The BiFCA result shows that 100% of extracted biclusters are statistically significant with adjusted *p-value*, $p < 0.001\%$. Worth of mention the best of the other compared algorithms is BiMine when $p < 0.1\%$, while CC, ISA and OSPM have reasonable performance with $p < 0.5\%$.



⁵ Available at <http://www.yeastgenome.org/go/GO:0002181/overview>

⁶ Available at <http://lama.mshri.on.ca/funcasssociate/>

⁷ The adjusted significance scores assess genes in each bicluster, which indicates how well they match with the different GO categories

⁴ Available at <http://db.yeastgenome.org/cgi-bin/GO/goTermFinder>

Proportions of Biclusters significantly enriched by GO annotations

V. CONCLUSION

In this paper, the BiFCA biclustering algorithm was presented as a new biclustering method for gene expression data sets.

Our approach relies on the extraction of formal concepts from a binary dataset. Given the nature of our initial dataset, we had to discretize it first into a -101 data matrix then into a binary data matrix.

The resulting formal concepts represent biclusters that we filter with the help of the measure bond.

The performance was evaluated with *Gene Ontology*(GO) annotations which checks the biological significance of biclusters.

Other avenues of future work concern the extraction of biclusters using generalized association rules, and we plan to apply BiFCA on different domains of application.

REFERENCES

- [1] Y. Cheng and G. M. Church, "Biclustering of expression data," in proc of ISMB, UC San Diego, California, 2000, pp. 93–103.
- [2] A. Freitas, W. Ayadi, M. Elloumi, L.-J. Oliveira, and J.-K. Hao, "Survey on biclustering of gene expression data," in *Biological Knowledge Discovery Handbook: Preprocessing, Mining, and Postprocessing of Biological Data*, 2013, pp. 591–608.
- [3] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE transactions on computational biology and bioinformatics*, vol. 1, pp. 24–45, 2004.
- [4] M. Kaytoue, S. O. Kuznetsov, A. Napoli, J. Macko, and W. M. Jr., "Caractérisation et extraction de biclusters de valeurs similaires avec l'analyse de concepts triadiques," in proc of EGC, 2012, pp. 65–76.
- [5] M. Kaytoue, S. O. Kuznetsov, and A. Napoli, "Biclustering numerical data in formal concept analysis," in proc of ICFA, Leuven, Belgium, 2011, pp. 135–150.
- [6] R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts." in *Ordered Sets*, I. Rival, Ed. Dordrecht/Boston: Reidel, 1982, pp. 445–470.
- [7] E. R. Omiecinski, "Alternative interest measures for mining associations in databases," *IEEE transactions on knowledge and data engineering*, vol. 15, pp. 57–69, 2003.
- [8] J. Hartigan, "Direct clustering of a data matrix," *Journal of the American Statistical Association*, vol. Volume 67, pp. 123–129, 1972.
- [9] A. Prelic, S. Bleuler, P. Zimmermann, A. Wille, P. Buhlmann, W. Gruissem, L. Hennig, L. Thiele, and E. Zitzler, "A systematic comparison and evaluation of biclustering methods for gene expression data," *Bioinformatics*, vol. 22, no. 9, pp. 1122–1129, 2006.
- [10] L. Teng and L. Chan, "Discovering biclusters by iteratively sorting with weighted correlation coefficient in gene expression data," *Journal of Signal Processing Systems*, vol. 50, pp. 267–280, 2008.
- [11] Y. Zhang, H. Zha, and C.-H. Chu, "A time-series biclustering algorithm for revealing co-regulated genes," in proc of The 5th International Conference on Information Technology: Coding and Computing, vol. 1, pp. 32–37, 2005.
- [12] K.-O. Cheng, N.-F. Law, W.-C. Siu, and A. W.-C. Liew, "Identification of coherent patterns in gene expression data using an efficient biclustering algorithm and parallel coordinate visualization," *BMC Bioinformatics*, vol. 9, no. 210, 2008.
- [13] W. Ayadi, M. Elloumi, and J.-K. Hao, "A biclustering algorithm based on a bicluster enumeration tree: application to dna microarray data," *BioData Mining*, vol. 2, p. 9, 2009.
- [14] J. Ihmels, S. Bergmann, and N. Barkai, "Defining transcription modules using large-scale gene expression data," *Bioinformatics*, vol. 20, pp. 1993–2003, 2004.
- [15] A. Tanay, R. Sharan, and R. Shamir, "Discovering statistically significant biclusters in gene expression data," *Bioinformatics*, vol. 18, pp. S136–S144, 2002.
- [16] W. Ayadi, M. Elloumi, and J.-K. Hao, "Bimine+: An efficient algorithm for discovering relevant biclusters of dna microarray data," *Knowl.-Based Syst.*, vol. 35, pp. 224–234, 2012.
- [17] S. Das and S. M. Idicula, "Application of cardinality based grasp to the biclustering of gene expression data," *International Journal of Computer Applications*, vol. 1, pp. 44–53, 2010.
- [18] W. Ayadi, M. Elloumi, and J.-K. Hao, "Iterated local search for biclustering of microarray data," in proc of PRIB, Nijmegen, The Netherlands, 2010, pp. 219–229.
- [19] F. Divina and J. S. Aguilar Ruiz, "Biclustering of expression data with evolutionary computation," *IEEE Transactions Knowledge and Data Engineering*, vol. 18, no. 5, pp. 590–602, 2006.
- [20] C. A. Gallo, J. A. Carballido, and I. Ponzoni, "Microarray biclustering: A novel memetic approach based on the pisa platform," in proc of EvoBIO, Tubingen, Germany, 2009, pp. 44–55.
- [21] S. Mitra and H. Banka, "Multi-objective evolutionary biclustering of gene expression data," *Pattern Recognition*, vol. 39, pp. 2464–2477, 2006.
- [22] K. C. Mondal, N. Pasquier, A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay, "A new approach for association rule mining and biclustering using formal concept analysis," in proc of MLDM, 2012, pp. 86–101.
- [23] R. G. Pensa, J. Besson, and J.-F. Boulicaut, "A methodology for biologically relevant pattern discovery from gene expression data," in *Discovery Science*, 2004, pp. 230–241.
- [24] F. Lehmann and R. Wille, "A triadic approach to formal concept analysis," *ICCS*, vol. 954, pp. 32–43, 1995.
- [25] Y. Luan and H. Li, "Clustering of time-course gene expression data using a mixed-effects model with bsplines," *Bioinformatics*, vol. 19, no. 4, pp. 474–482, 2003.
- [26] S. Peddada, E. Lobenhofer, L. Li, C. Afshari, and C. Weinberg, "Gene selection and clustering for time course and dose-response microarray experiments using order-restricted inference." *Bioinformatics*, vol. 19, pp. 834–841, 2003.
- [27] S. Ben Yahia, T. Hamrouni, and E. M. Nguifo, "Frequent closed itemset based algorithms: a thorough structural and analytical survey," *SIGKDD Explorations*, vol. 8, no. 1, pp. 93–104, 2006.
- [28] S. Tavazoie and J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, "Systematic determination of genetic network architecture genetics," *Nature genetics*, vol. 22, pp. 281–285, 1999.
- [29] Y. Cheng and G. M. Church, "Biclustering of expression data," supplementary information, Tech. Rep., 2006.
- [30] G. F. Berriz, O. D. King, B. Bryant, C. Sander, and F. P. Roth, "Characterizing gene sets with funcassociate," *Bioinformatics*, vol. 19, pp. 2502–2504, 2003.
- [31] S. Bergmann, J. Ihmels, and N. Barkai, "Defining transcription modules using large-scale gene expression data," *Bioinformatics*, vol. 20, no. 13, pp. 1993–2003, 2004.
- [32] A. Ben-Dor, B. Chor, R. M. Karp, and Z. Yakhini, "Discovering local structure in gene expression data: The order-preserving submatrix problem," *Journal of Computational Biology*, vol. 10, no. 3/4, pp. 373–384, 2003.
- [33] W. Ayadi, M. Elloumi, and J. K. Hao, "Bicfinder: a biclustering algorithm for microarray data analysis," *Knowl. Inf. Syst.*, vol. 30, no. 2, pp. 341–358, 2012.
- [34] W. Ayadi, "Algorithmes systématiques et stochastiques de biregroupement pour l'analyse des données biopuces," Ph.D. dissertation, University of Angers, France, 2011.

Simuler pour Evaluer les BMOs: Quelle Méthode Choisir ?

BADSI Hichem Benaïssa Anouar¹
GHOMARI Abdessamad Réda²
ZEMMOUCHI-GHOMARI Leila³

Laboratoire LMCS ,Ecole Nationale Supérieure d'Informatique (ESI)Alger, Algérie

¹h_bads@esi.dz

²a_ghomari@esi.dz

³l_zemmouchi@esi.dz

Abstract—Afin de tester une proposition d'indicateurs de performance destinés à mesurer l'impact des ontologies du modèle d'affaires (Business Model Ontologies : BMOs) sur les modèles d'affaires des entreprises, une approche de simulation s'avère indispensable. En effet, l'évaluation d'une ontologie avant et après son application dans le cadre d'une entreprise, mobiliserait une équipe d'experts sur une période qui reste à déterminer et sans garantie de résultats. Pour cela, il est primordial de sélectionner la méthode de simulation la plus adaptée à nos besoins.

Dans le cadre du présent travail de recherche, nous avons entrepris de réaliser un benchmarking des approches de simulation les plus populaires et de mettre leurs caractéristiques respectives à l'épreuve de nos contraintes.

Mots-Clés—Méthodes de simulation ; Evaluation par la Simulation ; Simulation en Entreprise ; Evaluation des ontologies des modèles d'affaires; Simulation à base d'Agents

I. INTRODUCTION

Le modèle d'affaires joue un rôle prépondérant dans l'amélioration du management de l'entreprise [1]. Il peut manifestement bénéficier de la caractéristique de formalité des ontologies [2]. Toutefois, la littérature relative à ce domaine manque de cas concrets et d'éléments mesurables qui permettraient l'évaluation de ce type d'ontologies sur les modèles d'affaires des entreprises [3].

Dans un précédent travail[4], nous avons proposé un ensemble d'indicateurs de performance ayant pour objectif de mesurer l'impact des ontologies des modèles d'affaires sur les entreprises. Ces indicateurs sont 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 modèle d'affaires (Business Model Ontology : BMO) dans une entreprise donnée. Ce type d'expérimentation comprend plusieurs contraintes, que nous énumérerons dans ce qui suit :

- la contribution nécessaire d'une équipe d'experts en modèles d'affaires, de développeurs d'ontologies

et de différentes parties prenantes de l'entreprise dans laquelle se déroule l'évaluation ;

- la durée de l'évaluation doit être déterminée de façon empirique même s'il paraît évident qu'elle s'étendra sur plusieurs exercices comptables afin d'obtenir des résultats probants ;
- la qualité des résultats obtenus n'est pas garantie (pas concluante), étant donné que ce type d'expérimentation n'a pas été réalisé à notre connaissance.

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.

En effet, la simulation par ordinateur est capable de fournir des résultats précis sur des grands volumes de données en un temps réduit comparativement à l'approche classique des expérimentations dans des conditions réelles d'exploitation [5].

Toutefois, la sélection de l'approche de la modélisation de la simulation la plus appropriée pour le cas d'étude est souvent tributaire de sa facilité d'utilisation aux dépens des critères d'adéquation ou d'efficacité[6]. Nous nous attellerons à réaliser une étude comparative objective entre les différents systèmes de simulation.

L'objectif de cet article est de réaliser un benchmarking des méthodes de simulation existantes, de les comparer et d'en sélectionner une, sur la base de critères dictés par les besoins et les contraintes de notre problématique de recherche qui est d'évaluer l'impact des BMOs sur les modèles d'affaires des entreprises.

Notre article est organisé de la manière suivante : en Section II nous expliquons l'importance des approches de simulation pour les entreprises. La Section III comprend la description et la comparaison des principales approches de simulation utilisées dans le milieu des entreprises. En Section IV, nous nous focalisons sur les cas d'utilisation de la simulation à base d'agents dans les entreprises à des fins d'évaluation. La Section V présente la synthèse de ce travail et sa conclusion.

II. SIMULATION DANS LE CONTEXTE D'UNE ENTREPRISE

En recherche, les méthodes d'évaluation sont diverses[7], elles peuvent être: descriptives, observationnelles, analytiques ou encore expérimentales.

L'évaluation expérimentale est divisée en deux catégories:

- l'expérimentation contrôlée, qui revient à étudier l'artefact à évaluer dans un environnement contrôlé selon certains critères comme l'utilisabilité ;
- ou la simulation qui teste l'artefact avec des données artificielles.

Par ailleurs, la simulation permet aux chercheurs d'assumer la complexité inhérente aux systèmes d'organisation comme une donnée, et par conséquent permet l'étude de systèmes plus complexes [8]. Elle a la particularité de générer des observations en se projetant dans le futur, alors que les autres méthodes d'évaluation considèrent l'historique pour déduire des conclusions.

De plus, la nature des systèmes vivants implique l'augmentation de l'entropie (désordre) ou de la complexité (l'ordre), ainsi regarder vers l'arrière est intrinsèquement plus difficile que vers l'avant [8].

Depuis sa création, la simulation a été appliquée à divers secteurs, tels que l'industrie, les services, la défense, la santé et les services publics. Elle est reconnue comme l'une des techniques les plus largement utilisées dans le domaine de la gestion des opérations (supervision de la conception et le contrôle du processus de production). Son utilisation a été transformée par l'invention et l'évolution de l'ordinateur, qui a appuyé l'adoption des outils et des techniques pratiques de simulation [9].

Ainsi selon Liotta[6], la pertinence de la simulation pour les entreprises a fait ses preuves en tant qu'outil pour la conception et la gestion des réseaux de chaîne d'approvisionnement complexes ou des systèmes de production.

La Figure 1 illustre bien le principe de simulation comme moyen de résoudre un problème en passant par l'alignement de la réalité avec un modèle abstrait. Il y a une distinction entre le modèle analytique et le modèle de simulation qui apporte un atout majeur dans la possibilité de multiplier les itérations à volonté (jusqu'à obtention d'un modèle optimisé); contrairement au modèle analytique où le modèle peut être inexistant ou extrêmement difficile à trouver [9].

Nous avons ajouté à la Figure 1 des indicateurs de performance qui pourront être mieux analysés et évalués

dans le modèle abstrait, et ainsi par la simulation être optimisés et mappés à nouveau vers la réalité. Ce principe est valable pour tous les types d'indicateurs y compris ceux que nous avons sélectionnés pour l'évaluation de l'impact des BMO en entreprise. Par exemple dans la Figure 1, nous considérons comme indicateurs ambigus ceux utilisés par l'entreprise sans avoir été validés. Le passage à un modèle abstrait simulable permettrait donc de confirmer ou d'infirmer la pertinence des indicateurs utilisés (erroné, moyen, idéal).

Concrètement, le recours à la simulation, en tant que méthode d'évaluation, comporte les avantages suivants[10, 11]:

- Effectuer des tests à moindre coût avant l'adoption de la technologie ;
- Multiplier les tests, en termes de type et de nombre, dans des environnements différents et sur une période relativement réduite;
- Evaluer avec une plus grande précision l'impact des technologies adoptées sur la performance des entreprises ;
- Effectuer des prévisions à long terme en simulant l'évolution de l'entreprise dans le temps ;
- Obtenir des résultats homogènes et facilement comparables qui faciliteront la prise de décision.

Cependant, des obstacles freinent l'adoption à grande échelle de la simulation distribuée et parallèle dans le contexte des entreprises tels que, le manque d'efficacité et de simplicité d'utilisation des systèmes de simulation et le manque de puissance du matériel utilisé.

En effet, pour que les outils de simulation soient adoptés par les décideurs des entreprises, il est indispensable de disposer d'outils efficaces qui produisent des modèles valides. Des architectures de systèmes de simulation facilitant l'interopérabilité et l'utilisation combinée des techniques d'optimisation et de simulation sont absolument nécessaires. Bien souvent, la vitesse d'élaboration des ordinateurs courants ne permet pas de réaliser des tests rapides de modèles d'optimisation complexes en association avec la simulation [6]

De ce fait l'adéquation et la pertinence de techniques de simulation est un facteur clé dans la réussite des applications pratiques en entreprise car il existe un besoin croissant d'aborder la complexité de l'ensemble de l'entreprise comme un tout d'une part, et d'autre part, la difficulté de traiter cette complexité avec les différents niveaux de prise de décision[9].

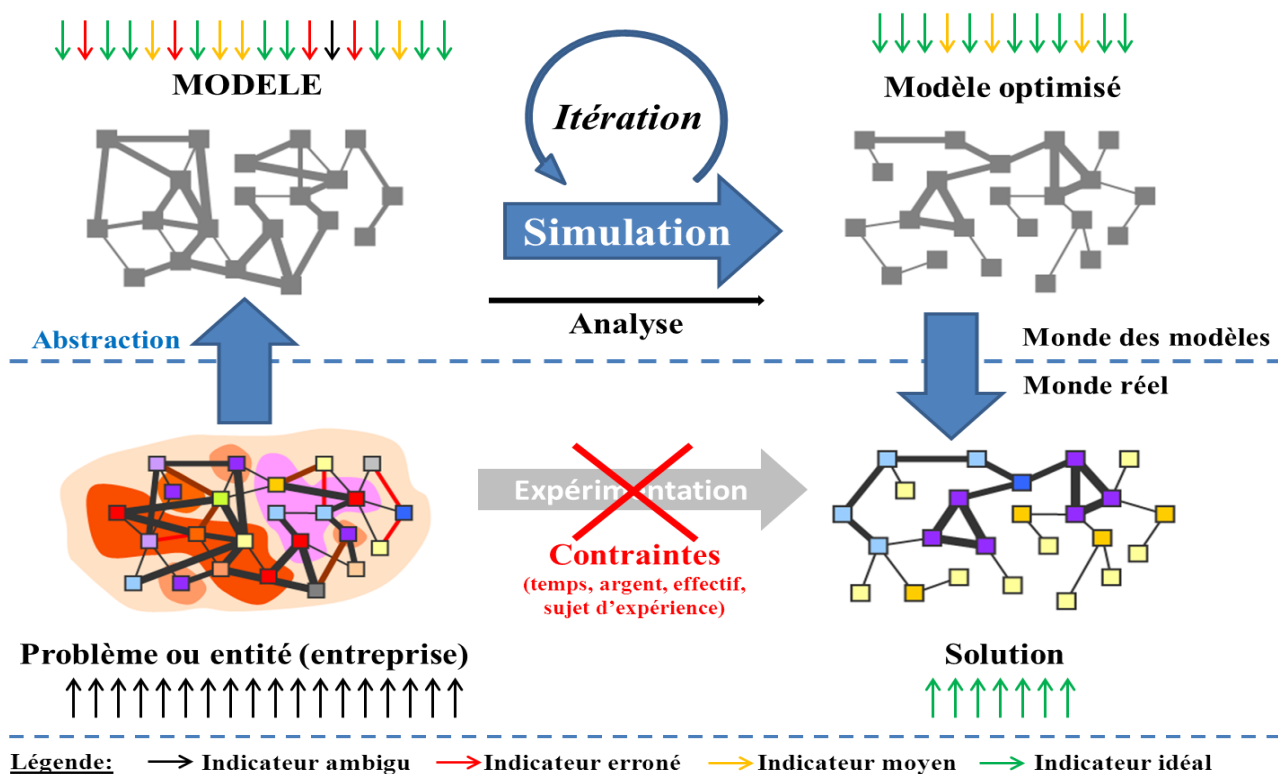


Figure 1. Modélisation de la simulation des indicateurs, Adapté de [12]

III. COMPARAISON DES METHODES DE SIMULATION

Jahangirian [9] a réalisé une étude sur le nombre d'applications de la simulation par domaine d'étude. D'après cette recherche la *Planification* est le domaine d'application le plus référencé, ce qui concorde avec les résultats rapportés dans [13]. Le génie des procédés dans l'industrie manufacturière, qui comprend à la fois la conception et l'amélioration des processus, représente la deuxième application la plus populaire de la simulation, tandis que d'autres applications, comme la gestion de la chaîne de logistique, la stratégie, le Transport et la Gestion de projet arrivent plus bas dans la liste.

Ces mêmes travaux ont relevé une augmentation significative de la modélisation de la stratégie par la simulation, ce qui est probablement dû à l'utilisation croissante de la dynamique des systèmes. En outre, le nombre de documents sur la *gestion de projet*, la *formation à la gestion*, la gestion des connaissances et la conception organisationnelle est non négligeable.

Cinq techniques de simulation ont été répertoriées comme étant les plus utilisées dans les secteurs du business et de l'industrie [9], notamment :

- La simulation à événements discrets (Discrete Event Simulation, DES)
- La dynamique des systèmes (System Dynamics, SD)
- La simulation à base d'agents (Agent Based Systems, ABS)
- Les jeux de simulation (Simulation Games, SG)
- La simulation hybride (Combinaison de SD et de DES)

D'autres approches de simulation méritent d'être citées telles que : la simulation de trafic, la simulation distribuée, la simulation virtuelle, les réseaux de Pétri, la simulation de Monte-Carlo et la simulation intelligente.

La littérature est prolifique en termes d'articles décrivant les différences entre ces systèmes de simulation [5, 8, 12, 14, 15]. Les approches de simulation étant nombreuses et variées, nous nous focaliserons sur les trois principales : DES, SD et ABS.

Dans ce qui suit, nous allons définir brièvement chacune des trois approches de simulation présélectionnées avant d'aborder l'étude comparative.

La *simulation à événements discrets (DES)* [16] est historiquement la plus commune, celle-ci suppose que le système organisationnel peut être décrit comme une « machine » où les intrants (entités) arrivent et sont transformés (par des événements) en sorties (entités). L'incertitude associée à la *vraie vie* est mise en œuvre sous la forme de variables aléatoires. La simulation à événements discrets est bonne pour étudier les éléments les plus mécaniques, prévisibles, et ordonnés de l'organisation [8].

La *dynamique des systèmes (SD)* [17] traite les systèmes organisationnels quelque peu mécaniques, et peut-être même déterministes. Dans la mesure où l'on peut décrire une organisation par ses États constitutifs et leurs interrelations en termes de taux ; la dynamique du système offre un riche paradigme de modélisation. Les entreprises utilisent souvent des modèles de dynamique du système à des fins de planification stratégique, mais ces efforts voient rarement le jour aux yeux du public. Selon Dooley [8], il est préférable d'utiliser les modèles de dynamique des systèmes à des fins spécifiques car son utilité sur les

problèmes génériques est marginale. Par exemple, si l'objectif visé est «l'apprentissage organisationnel», de meilleurs résultats découleraient de l'observation d'une véritable organisation.

Les modèles à base d'agents (ABS) [18], se concentrent sur les membres de l'organisation et prévoient que le comportement organisationnel émerge des interactions couplées des membres de l'organisation. Ces modèles né de l'intelligence artificielle ont fusionné avec la science de la complexité pour offrir un paysage plus complet et plus solide théoriquement au sein duquel, il est possible de développer des modèles intéressants et instructifs. Les modèles à base d'agents sont probablement les mieux adaptés pour répondre aux questions typiques des chercheurs organisationnels [8].

D'un point de vue théorique, l'ensemble des modèles de SD sont un sous-ensemble strict de l'ensemble de tous les modèles ABS [14]. C'est le théorème d'agence (Agency Theorem) pour les dynamiques des systèmes, qui déclare que chaque modèle SD bien formulé a une formulation équivalente dans un modèle ABS. Par conséquent, il est possible de modéliser n'importe quel modèle SD en utilisant le modèle ABS. En déduction, il est possible d'obtenir des résultats au moins aussi bons que par l'exécution d'un modèle SD même si les modèles ABS consomment plus de temps en termes de modélisation et d'exécution [14].

D'autres travaux ont comparé les systèmes ABS avec les systèmes traditionnels DES [15] en mettant en évidence leurs différences respectives. Tout d'abord dans les systèmes ABS, les agents ont chacun leur propre comportement et sont donc considérés comme « Actifs ». En revanche, les comportements des entités dans les modèles DES sont déterminés par les systèmes et sont, de ce fait, classés « passifs ». Une autre différence clé est que le modèle DES est construit autour des réseaux de files d'attente, alors que dans un système ABS, il n'y a pas de notion de files d'attente.

Malgré leurs différences, DES et ABS partagent certaines similitudes. Les deux sont de nature stochastique et peuvent utiliser les distributions en entrée pour modéliser un comportement aléatoire [14]. Il est également possible de les combiner par la création d'un système DES, dans lequel certaines entités sont, en fait, actives.

ABS a souvent été négligée au profit de méthodes traditionnelles (DES et SD), cela est dû à plusieurs raisons. Tout d'abord, les logiciels disponibles sont assez difficiles à utiliser et donc exclus par un grand nombre de modélisateurs. De plus, les modèles ABS sont exigeants en termes de temps de développement. Enfin, nous retiendrons la réticence des praticiens de la recherche opérationnelle envers la modélisation à base d'agents au profit de la modélisation à événements discrets ou la dynamique des systèmes [14].

Notre propre étude comparative entre les systèmes de simulation, précédemment cités, est présentée dans un tableau de synthèse (Tableau 1), et ce sur la base de plusieurs critères définis dans [5, 8, 12, 14, 15], à savoir:

Niveau d'abstraction : qui comprend trois niveaux différenciés par le niveau de granularité (détail) croissant

des entités prises en considération (par exemple : entreprise, service ou personne), à savoir : le niveau stratégique, le niveau tactique et le niveau opérationnel.

Processus : Les modèles continus peuvent représenter des modifications à des moments précis dans le temps, alors que dans les modèles discrets, la progression du temps se fait par unités fixes.

Nature : Les modèles déterministes présument que tous les paramètres du modèle sont connus et fixés avec certitude, alors que les modèles stochastiques tiennent compte de paramètres incertains et aléatoires.

Approche : « top-down » où l'accent est mis sur la modélisation du système, pas les entités, alors qu'en « bottom-up », l'accent est mis sur la modélisation des entités et des interactions entre elles.

Objets : un objet « Actif » veut dire que chaque entité a son propre comportement alors que pour un objet « passif » le comportement est déterminé par le système.

Entités : la capacité du modèle à supporter un grand nombre d'entité et leur hétérogénéité.

D'autres critères de comparaison concernent la décentralisation et la capacité du modèle à représenter des modèles complexes, l'environnement et les comportements individuels, ainsi que la facilité de modification des processus.

Critère/Valeur		DES	SD	ABS
Niveau d'abstraction	Stratégique		X	X
	Tactique	X		X
	Opérationnel	X		X
Processus	Discret	X		X
	Continu		X	
Nature	Déterministe		X	
	Stochastique	X		X
Approche	Top-down	X	X	
	Bottom-up			X
Objets	Actif			X
	passif	X	X	
Entités	Nombre élevé		X	X
	Hétérogénéité	X		X
Environnement		X		X
Modélisation des comportements individuels				X
Décentralisation				X
Facilité de modifications des processus				X
Prise en charge de modèles complexes				X

Tableau 1. Comparatif des méthodes de simulation

Le tableau comparatif met en évidence la quasi-complétude de la prise en charge des critères prédéfinis par la modélisation à base d'agents.

Notre analyse de la problématique a mis en évidence les avantages apportés par les systèmes ABS ainsi que les similarités entre la systémique des entreprises et celle des agents, ce qui nous amène naturellement à considérer cette technique comme la plus adéquate au regard de nos objectifs. Cette déduction est confortée par les résultats des cas d'utilisation d'ABS en entreprises identifiés dans la littérature que nous détaillons dans la Section suivante.

IV. SIMULATION A BASE D'AGENTS DANS LE CONTEXTE D'UNE ENTREPRISE

Des cas intéressants d'utilisation de la simulation à base d'agents (ABS) pour l'entreprise ont été cités dans la littérature. Nous en décrivons une sélection dans ce qui suit, afin d'illustrer la faisabilité et la rentabilité de ce type de simulation dans le contexte des entreprises pour des besoins d'évaluation a priori:

Dans [19], les auteurs proposent un modèle évolutif de réseau d'approvisionnement afin de comprendre le principe général de son évolution. Pour cela, ils effectuent une simulation multi-agents sur le modèle, et montrent que le réseau d'approvisionnement émerge et évolue de l'interaction dynamique des entreprises dans un environnement dynamique, dominé par l'environnement et les mécanismes internes des entreprises. Toutefois, l'évolution est très sensible par rapport à l'état initial et est difficile à prévoir avec précision.

Dans [20], les auteurs ont utilisé le modèle ABM pour développer un modèle de marché de la consommation multi-échelle. L'objectif du modèle obtenu est de tester la robustesse des stratégies de marché et de permettre une exploration des facteurs de tendances potentiels. Ils ont ensuite effectué des tests d'étalonnage, de vérification et de validation de ce modèle. Selon eux, le modèle a été appliqué avec succès par l'entreprise Procter & Gamble, à plusieurs problèmes d'affaires complexes, dans lesquels, le modèle a directement influencé la prise de décision managériale et produit des économies substantielles.

Dans [21], Les auteurs ont formalisé et paramétré un modèle multi-agents détaillé pour la production de six carburants et six mélanges combustibles de six matières premières à travers 13 chaînes de production, et leur adoption par 11 sous-populations distinctes d'automobilistes. Ce modèle améliore les modèles traditionnels d'optimisation des coûts en comprenant des attributs non-économiques et des commentaires sociaux. L'adoption de carburants de remplacement étant le plus souvent limitée à des créneaux de marché dépassant pas les 5%. Un seul cas de transition complet de carburant a été trouvé. La variété des résultats indique qu'il est impossible de faire des gains grâce à des technologies de carburant alternatives.

Dans [22], les auteurs considèrent l'ABM comme un outil viable pour les auditeurs, permettant de tenir compte de l'interaction de facteurs locaux et environnementaux afin de déterminer le succès organisationnel. Pour illustrer cela, ils ont développé un ABM du marché des consommateurs du transport aérien pour la compagnie Frontier Airlines en 2007. Le modèle génère un pourcentage de part de marché prévisionnel pour l'année suivante. Après de

nombreuses simulations, le système a généré une plus-value de part de marché supérieure de l'ordre de 0,17%, par rapport au chiffre réel obtenu par Frontier en 2007.

Dans [23], les auteurs combinent la technologie multi-agent avec des solutions basées-ontologie pour construire un système sémantique d'aide à la décision, nommé SemDSS, qui est en mesure de soutenir et d'automatiser en partie les processus d'approvisionnement des marchés. Les auteurs introduisent des fondements sémantiques dans le but d'améliorer le traitement automatisé des données dans les environnements sémantiquement hétérogènes. Leur approche propose des solutions pour résoudre le problème de l'interopérabilité dans le processus d'approvisionnement.

Dans [24], les auteurs étudient la dynamique collective d'un réseau de chaîne d'approvisionnement, utilisant une approche de modélisation à base d'agents. Ils ont introduit des éléments stochastiques au modèle et observé comment les événements locaux affectent la performance globale du système. Leur modèle modifié représente, plus en détail, l'environnement économique réel dans lequel opèrent les entreprises. Et cela afin de simuler l'environnement économique des entreprises de façon réaliste, en ajoutant les fonctions suivantes: l'évolution d'un réseau de chaîne d'approvisionnement avec la reconfiguration de liens, la dispersion des prix, la dynamique des prix et les coûts de production.

Dans [25], les auteurs montrent qu'il est possible de générer des journaux d'événements à partir de simulations multi-agents. Les interactions entre les agents peuvent être enregistrées grâce à leur plateforme afin de les utiliser pour l'analyse et l'extraction de processus. Grâce à la description de haut niveau du processus d'affaires, ce journal des événements peut être utilisé pour extraire un nouveau modèle hiérarchique ou analyser et évaluer ce modèle afin de cerner d'autres possibilités de changements. Ce cycle d'amélioration peut être répété jusqu'à ce qu'un modèle de processus soit satisfaisant.

Dans [26], les auteurs tentent de révéler les facteurs clés pour l'évolution des grappes industrielles dans le cadre du développement dynamique de l'environnement extérieur et de l'innovation interne. Pour ce faire un modèle de système multi-agent est construit et l'algorithme d'apprentissage du modèle est basé sur un algorithme génétique. L'étude de cette méthode montre que l'évolution des grappes industrielles provient de l'interaction complexe des agents internes et que les principales actions des entreprises prennent des initiatives sur les facteurs fondamentaux dans le processus d'évolution des grappes industrielles.

Le tableau croise les cas d'utilisation décrits précédemment avec quelques indicateurs proposés dans [4] pour évaluer l'impact des BMO dans le contexte d'une entreprise.

Nous utilisons le symbole (+) ou (-) pour illustrer l'existence ou l'absence de la modélisation des indicateurs concernés dans les cas cités. Dans le cas favorable (+), il est possible de déduire que l'indicateur en question est

évaluable par un modèle ABS. Le symbole (/), signifie que l'information recherchée est indisponible.

Indicateurs/ Cas d'Utilisation	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]
Quel référentiel de qualité est adopté pour le développement d'un service et/ou d'un produit ?	+	/	/	/	+	/	/	/
Quelle est la capacité de production ?	+	/	+	-	/	+	/	+
Quel est le taux de rejet de produits ?	/	/	/	/	/	+	/	/
Quel est le temps moyen de fabrication de produits ?	+	/	-	/	-	-	/	/
Quel est le volume de Vente / d'Achat ?	/	+	+	+	+	/	/	/
Quel est le nombre de clients ?	/	+	+	+	+	/	/	/
Quel est le bénéfice généré par un client ou une catégorie de clients ?	/	+	+	+	+	/	/	/
Est-ce que l'entreprise entretient des relations durables avec ses clients ?	/	-	-	+	/	-	/	/
Quelle est le nombre de clients attirés ou acquis ?	/	+	+	+	/	-	/	/
Quel est le niveau d'autonomie de l'employé ?	/	/	/	/	/	/	/	/
De quelle(s) formation(s) a bénéficié l'employé ?	/	/	/	/	/	/	/	/

Tableau 2. Cas d'utilisation vs. Indicateurs

V. SYNTHÈSE ET CONCLUSION

L'étude des systèmes de simulation existants et leurs caractéristiques respectives est un travail incontournable pour réaliser une sélection objective et fondée pour répondre aux besoins de l'évaluation des BMOs dans le contexte des entreprises.

En se basant sur la littérature, sur notre étude comparative (Section III) et sur les différents cas rapportés en Section IV, il en résulte que :

En premier lieu, la structure d'une modélisation à base d'agents, est plus en accord avec celle de l'entreprise. Puisqu'on retrouve la communication et l'interaction entre agents humains et agents artificiels pour atteindre des objectifs individuels ou collectifs. Ces agents, peuvent se spécialiser et/ou former des équipes pour résoudre un problème donné ou atteindre un objectif global de l'organisation. Notons également, l'équivalence entre l'interdépendance et l'interaction des agents avec leur environnement et celle des employés avec leur environnement de travail direct ou avec les entreprises faisant partie de leur environnement économique, politique ou géographique.

Dans un second lieu cette concordance entre les systèmes à base d'agents et l'entreprise permet une meilleure compréhension des phénomènes inhérents à ces organisations. En effet, il sera plus simple de suivre le déroulement d'événements sur un modèle fonctionnant avec la même logique que la réalité, où l'imprévisible est généré par l'interaction et non par des variables aléatoires comme dans les systèmes de simulation à événements discrets.

En troisième lieu, Le démontre bien qu'une modélisation à base d'agents permet indubitablement de modéliser quelques indicateurs de performance définis dans nos travaux précédents [4], et que leur simulation par ce biais nous permettra de valider leur pertinence pour la mesure de la performance des BMOs .

On remarque dans le tableau ci-dessus, que la plupart des entités relatives à nos indicateurs sont modélisables excepté pour les indicateurs relatifs aux employés qui ne sont pas pris en charge par les cas cités dans la section précédente. Toutefois, la modélisation à base d'agents est bien connue pour ses aptitudes à modéliser les comportements sociaux [27].

De plus, le modèle de simulation que nous prospectons se doit d'être le plus générique possible (non spécifique ou relatif à une catégorie d'entreprise), par conséquent, la simulation par la dynamique des systèmes n'est donc pas recommandée pour répondre à nos besoins.

D'ailleurs, Borshchev et Filippov [12] affirment que la modélisation à base d'agents est destinée à ceux qui souhaitent aller au-delà des limites des approches SD et DES, en particulier dans les cas où le système modélisé contient des objets actifs (personnes, unités d'affaires, animaux, véhicules, projets, et produits) ainsi que la synchronisation, la commande d'événements ou d'autres types de comportement individuel.

De ce fait, des trois modèles présentés et comparés, la modélisation à base d'agents s'impose naturellement pour répondre aux besoins de notre évaluation, et cela malgré son principal inconvénient: la dualité complexité/fonctionnalité des outils y afférents.

Par conséquent, l'approche la plus prometteuse pour la suite de nos travaux de recherche s'avère être la modélisation à base d'agents, qui par ses similitudes naturelles avec l'entreprise devrait nous apporter le plus d'information pertinentes pour confirmer ou infirmer nos hypothèses de départ.

Les perspectives de nos travaux s'inscrivent dans l'élaboration d'un modèle de simulation multi-agent avec lequel il sera hypothétiquement possible d'évaluer l'impact des ontologies du business model sur les modèles d'affaires des entreprises

REFERENCES

- [1] Osterwalder, A., *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, 2004. 173.
- [2] Dietz, J.L., System Ontology and its role in Software Development. EMOI-INTEROP, 2005. 5.
- [3] Zott, C., R. Amit, and L. Massa, *The business model: recent developments and future research*. Journal of management, 2011. 37(4): p. 1019-1042.
- [4] Badsì, H.B.A., 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), xxxx. x(x): p. (in press).
- [5] Scholl, H.J. *Agent-based and system dynamics modeling: a call for cross study and joint research*. in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*. 2001. IEEE.
- [6] Liotta, G., *Simulation of supply-chain networks: a source of innovation and competitive advantage for small and medium-sized enterprises*. Technology Innovation Management Review, 2012(November: Insights).
- [7] Hevner, A.R., S.T. March, J. Park, S. Ram, *Design science in information systems research*. MIS Quarterly, 2004. 28(1): p. 75–105.
- [8] Dooley, K., *Simulation research methods*. Companion to organizations, 2002: p. 829-848.
- [9] Jahangirian, M., T. Eldabi, A. Naseer, L.K. Stergioulas, T. Young, *Simulation in manufacturing and business: A review*. European Journal of Operational Research, 2010. 203(1): p. 1-13.
- [10] Rejeb, L., *simulation multi-agents de modèles économiques Vers des systèmes multi-agents adaptatifs*. 2005, Thèse de doctorat de l'Université Reims Champagne-Ardennes.
- [11] Lopez-Sanchez, M., X. Noria, J.A. Rodríguez, N. Gilbert, *Multi-agent based simulation of news digital markets*. International Journal of Computer Science & Applications, 2005. 2(1): p. 7-14.
- [12] Borshchev, A. and A. Filippov. *From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools*. in *Proceedings of the 22nd international conference of the system dynamics society*. 2004.
- [13] Shafer, S.M. and T.L. Smunt, *Empirical simulation studies in operations management: context, trends, and research opportunities*. Journal of Operations Management, 2004. 22(4): p. 345-354.
- [14] Maidstone, R., *Discrete Event Simulation, System Dynamics and Agent Based Simulation: Discussion and Comparison*. System, 2012: p. 1-6.
- [15] Siebers, P.-O., C.M. Macal, J. Garnett, D. Buxton, M. Pidd, *Discrete-event simulation is dead, long live agent-based simulation!* Journal of Simulation, 2010. 4(3): p. 204-210.
- [16] Law, A.M., W.D. Kelton, and W.D. Kelton, *Simulation modeling and analysis*. Vol. 2. 1991: McGraw-Hill New York.
- [17] Forrester, J.W., *Industrial dynamics*. Vol. 2. 1961: MIT press Cambridge, MA.
- [18] Macal, C.M. and M.J. North, *Tutorial on agent-based modelling and simulation*. Journal of Simulation, 2010. 4(3): p. 151-162.
- [19] Li, G., P. Ji, L.Y. Sun, W.B. Lee, *Modeling and simulation of supply network evolution based on complex adaptive system and fitness landscape*. Computers & Industrial Engineering, 2009. 56(3): p. 839-853.
- [20] North, M.J., C.M. Macal, J.st Aubin, P. Thimmapuram, M. Bragen, J. Hahn, J. Karr, N. Brigham, D. Hampton, *Multiscale agent-based consumer market modeling*. Complexity, 2010. 15(5): p. 37-47.
- [21] Vliet, O.v., B. Vries, A. Faaij, W. Turkenburg, W. Jager, *Multi-agent simulation of adoption of alternative fuels*. Transportation Research Part D: Transport and Environment, 2010. 15(6): p. 326-342.
- [22] Kuhn, J.R., J.F. Courtney, and B.J. Morris, *Agent Based Simulation and the Auditor Going Concern Opinion: A Case Study of Frontier Airlines*. International Journal of Business and Social Science, 2011. 2(22): p. 45-51.
- [23] Gottfried, K., M. Merdan, W. Lepuschitz, T. Moser, C. Reinprecht, *Multi Agent Systems combined with Semantic Technologies for Automated Negotiation in Virtual Enterprises*. Multi-Agent Systems-Modelling, Control, Programming, Simulations and Applications, 2011: p. 221-242.
- [24] Mizgier, K.J., S.M. Wagner, and J.A. Holyst, *Modeling defaults of companies in multi-stage supply chain networks*. International Journal of Production Economics, 2012. 135(1): p. 14-23.
- [25] Ferreira, D.R., F. Szimanski, and C.G. Ralha, *Mining the low-level behaviour of agents in high-level business processes*. International Journal of Business Process Integration and Management, 2013. 6(2): p. 146-166.
- [26] Yang, Y. and W. Niu, *Multi-Agents Model and Simulation for the Evolution of Industrial Clusters*. Journal of computers, 2013. 8(2).
- [27] Bonabeau, E., *Agent-based modeling: Methods and techniques for simulating human systems*. Proceedings of the National Academy of Sciences of the United States of America, 2002. 99(Suppl 3): p. 7280-7287.

Exploration diversifiée par entités nommées

Imène Saidi Université d'Oran,
Laboratoire LITIO
BP 1524, El-M'Naouer, 31000
Oran, Algérie

Sihem Amer Yahia
CNRS, LIG
Grenoble, France

Safia Nait Bahloul Université
d'Oran, Laboratoire LITIO
BP 1524, El-M'Naouer, 31000
Oran, Algérie

Abstract—Le web est devenu la principale source où les utilisateurs trouvent et publient des informations. Les pages web portent souvent sur un sujet particulier et sont écrites autour d'entités nommées (une personne, un pays, etc..) mais quelques moteurs de recherche restent inadéquats à certains types de recherche. En effet, ce que les utilisateurs cherchent réellement dans le web ne sont pas les pages web mais des informations que celles-ci contiennent (dans ce cas, des entités). Dans cet article, nous considérons le problème de la recherche d'entités pertinentes relatives aux requêtes des utilisateurs. Les entités recherchées pouvant être connues ou inconnues, les utilisateurs devraient avoir le choix de poser leurs requêtes de différentes manières (recherche par une/des entités, recherche par mots clés relatifs). Nous considérons également la multitude de types d'entités (Paris : ville, personne) et de catégories des documents contenant les entités (Politique, Sports, ...). Nous proposons alors une approche qui traite différents types de requêtes (entités connues, ou inconnues décrites par des mots dont l'interprétation sera faite autour d'entités) et permet l'exploration des résultats des entités trouvées en les diversifiant par les types et les catégories identifiés. Nos expériences démontrent l'efficacité de notre approche.

Keywords— Recherche d'entités; diversité des résultats de la recherche; Recherche d'informations; indexation

I. INTRODUCTION

La recherche d'information suscite depuis fort longtemps l'attention de la communauté scientifique mais avec l'expansion d'internet, la mise en œuvre de solutions capables d'exploiter le contenu du web et d'améliorer la performance de la recherche est devenue primordiale. Des techniques ont été proposées et des applications ont été réalisées, leur objectif est de fournir aux utilisateurs des réponses pertinentes par rapport aux besoins qu'ils expriment. Dans certains cas, ce que l'utilisateur cherche réellement dans le web n'est pas les pages web, mais des informations que celles-ci contiennent (c.-à-d. des unités ou des entités) d'où l'idée de la recherche directe d'entités [1]. Trouver des entités à la place des documents dans le web est un axe de recherche récent dans la recherche d'informations (RI). Les travaux de l'état de l'art motivent leurs approches de la recherche d'entités dans le contexte du web. Dans notre travail, nous nous intéressons à des domaines d'applications spécifiques où la plupart des documents sont écrits autour d'entités nommées et sont organisés par thèmes. Parmi ces domaines, nous pouvons trouver les forums de discussion, les articles de journaux, les wiki news, etc.

Dans ce travail, nous construisons un système de recherche d'entités qui prend en charge des requêtes construites par une ou plusieurs entités (Recherche par entités) ainsi que les requêtes de mots clés (Recherche

d'entités par des mots clés) et retourne des entités relatives aux requêtes.

Dans notre travail, nous visons à offrir à l'utilisateur la possibilité de trouver des entités pertinentes aux différentes requêtes tout en augmentant la diversité des résultats. Le problème de la diversité des résultats de la recherche a été traité dans de nombreux travaux dans le contexte de recherche de documents [2] [3] [4], néanmoins, la définition que nous allons conférer à la diversité nous permettra de contourner sa complexité qui a été définie comme un problème NP-complet.

Notre définition de la diversité se base sur différentes caractéristiques des entités, à savoir, les types des entités (personne, pays, ...) et les catégories des documents qui contiennent les entités (politique, sport, ...). A notre connaissance, ce type de diversité est nouveau et n'a jamais été proposé dans les travaux de l'état de l'art qui sont relatifs à la recherche d'entités.

Notre approche permet l'exploration des différentes entités trouvées ainsi que leurs documents diversifiés (les documents portant sur chaque entité). Nous évaluons notre approche en utilisant le corpus 20NewsGroups⁸. Nos expériences démontrent l'efficacité de notre approche à retourner des résultats diversifiés aux utilisateurs.

Cet article est organisé comme suit. Dans la section 2, nous présentons quelques travaux relatifs à la recherche d'entités. La section 3 est consacrée à la présentation des exemples de motivation et la Section 4 à la solution du problème ainsi qu'à la présentation de l'algorithme. La section 5 est réservée aux expérimentations. Dans la section 6, nous présentons une conclusion et des perspectives.

II. TRAVAUX RELATIFS A LA RECHERCHE D'ENTITES

Le concept de la recherche d'entités a été présenté en [1] il a pour but d'exploiter la richesse du web afin d'en tirer les données enfouies dans les pages non structurées. Les premiers travaux [5] [6] [7] sur la recherche d'entités proposent principalement des approches efficaces pour le passage à l'échelle dans le contexte du web mais ne s'occupent pas de l'efficacité de la recherche des entités [8]. Dans [8], une architecture est proposée pour la recherche d'entités et des algorithmes efficaces sont présentés pour le classement des entités trouvées. Dans [9], les auteurs ont proposé un système pour une recherche d'entités efficace en se basant sur les tables relationnelles et dans [10], ils ont proposé une méthode pour faciliter les traitements nécessaires de la recherche en s'inspirant du succès que l'index inversé a connu dans la RI. La proposition est de concevoir des index qui facilitent le traitement en

⁸<http://qwone.com/jason/20Newsgroups/>

considérant les entités comme mots clés pour l'index. Dans notre travail nous nous baserons également sur la construction d'index qui nous sera utile pour le traitement des différentes requêtes.

Quelques approches de la recherche d'entités ont été proposées dans le contexte de Wikipédia, leur but est de classer les entités trouvées en exploitant les liens entre les pages de Wikipédia [11]. D'autres approches [12] [13] utilisent la sémantique et les techniques NLP (*Neuro-linguistic programming*) pour améliorer l'efficacité de la recherche d'entités. Dans un autre travail [14], le premier moteur de recherche d'entités a été présenté. Le système de recherche d'entités proposé crée un document de concordance pour chaque entité, ce document représente toutes les phrases contenant cette entité dans tout le corpus. Ces documents seront alors indexés et recherchés en utilisant des logiciels de recherche open-source, le résultat sera une liste des entités classées.

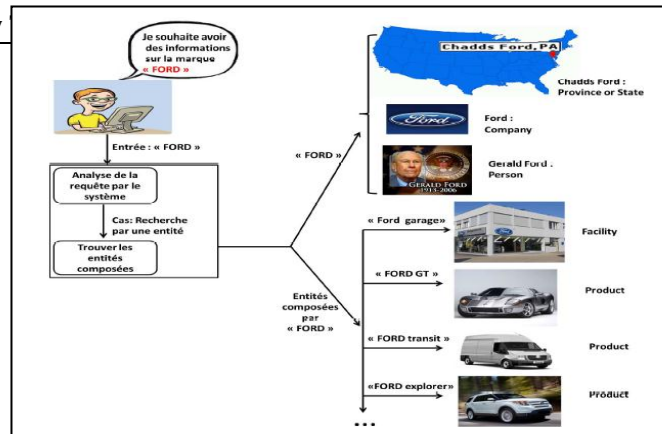
La recherche d'entités est également utilisée pour l'analyse des résultats des moteurs de recherche. Les auteurs de [15], proposent de chercher les entités dans les snippets⁹ des résultats, un snippet étant un texte de 10 à 20 mots. Ils proposent d'enrichir les systèmes de recherche classique par les résultats de la recherche des entités nommées. Cet enrichissement sera effectué au moment de traitement de la requête, aucun prétraitement n'est fait au préalable. L'inconvénient de cette approche est que les snippets ne contiennent pas toutes les entités nommées. Dans un autre travail [16], les auteurs étendent la recherche des entités du travail précédent [15] en considérant les textes complets des résultats. L'inconvénient de cette approche est le traitement des gros volumes de données et la consommation considérable des ressources, aucun prétraitement n'est effectué également.

Le graphe de connaissance "Google Knowledge Graph" [17], mis en place récemment par Google comprend les entités du monde réel et leurs relations et améliore la recherche de Google en retournant à l'utilisateur une description de l'entité qu'il cherche si la requête est composée d'une seule entité seulement.

La contribution de notre travail est de considérer différents types de requêtes et retourner des entités ainsi que leurs documents en interprétant les requêtes par les entités des documents de manière à maximiser la diversité des résultats. Dans les travaux cités précédemment, aucune diversification n'a été proposée pour les résultats.

III. EXEMPLES DE MOTIVATION

Dans le cadre de ce travail, les entités recherchées peuvent être connues lorsque l'utilisateur saisit une entité (Recherche par une Entité "RIE") ou plusieurs entités (Recherche par Plusieurs Entités "RPE") ou inconnues s'il ignore le nom de l'entité et saisit des mots clés relatifs (Recherche par Mots clés "RMC"). Supposons que l'utilisateur souhaite avoir des informations sur les marques de voitures. L'utilisateur peut poser différentes requêtes.



Exemple 1. Dans le premier scénario 1 (Fig. 1), l'utilisateur veut avoir des informations sur une marque de voiture particulière : ce cas représente la recherche d'une entité : RIE. L'utilisateur saisit l'entité (connue) qu'il veut trouver: "Ford", il aura en résultat l'entité "Ford" avec les entités composées par "Ford". L'entité Ford a plusieurs types, il peut s'agir de Gerald Ford l'ancien président d'Amérique, de la marque de voiture Ford ou encore d'une région en Amérique. Nous supposons que c'est plus intéressant de prendre en compte les différents types d'une même entité et de retourner à l'utilisateur les documents les plus pertinents de chaque type. Nous avons nommé cette approche : diversification par types. Les entités composées par "Ford" seront retournées aussi à l'utilisateur ainsi il pourra explorer leurs documents relatifs.

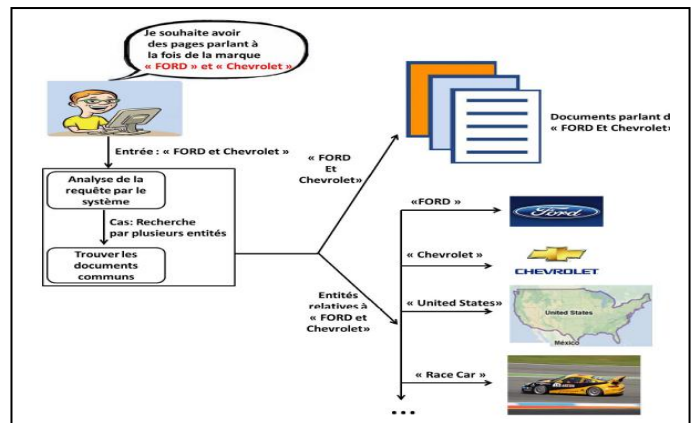


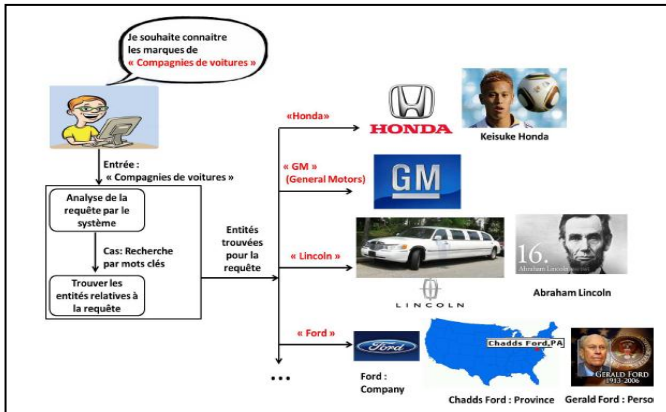
Fig. 2. Recherche par plusieurs entités "RPE" (Fig. 2),

l'utilisateur veut avoir des documents portant sur deux marques de voitures (il cherche un lien ou une comparaison entre les marques) : ce cas représente la recherche par plusieurs entités : RPE. L'utilisateur saisit les entités (connues) : "Ford et Chevrolet", il aura en résultat les documents portant sur les deux entités. Les entités relatives, qui apparaissent dans le même contexte seront retournées à l'utilisateur, nous les avons nommé : entités contextuelles. Nous supposons dans ce cas que c'est intéressant de prendre en compte les différentes catégories des documents les plus pertinents de la requête. Nous avons nommé cette approche : diversification par catégories. Les entités relatives seront retournées aussi à l'utilisateur ainsi il pourra explorer leurs documents relatifs.

Exemple 3. Dans le troisième scénario 3 (Fig. 3), l'utilisateur veut avoir des informations sur les différentes

⁹Le snippet est une description ou un extrait de la page qui apparaît sous son titre et son URL.

compagnies de voitures : ce cas représente la recherche d'entités par mots clés : RMC. L'utilisateur saisit sa requête : "compagnie de voiture", il aura en résultat les entités relatives (entités contextuelles) à cette requête et les résultats de chaque entité trouvée seront traités comme le cas de la recherche par entités (diversification par types si l'entité a plusieurs types sinon diversification par catégories).



IV. NOTRE APPROCHE

Pour certaines sources de données telles que les forums et les articles, nous proposons une recherche par mots-clés plus intéressante que la recherche par entités. Les entités existantes peuvent avoir plusieurs types et les documents les contenant peuvent avoir plusieurs catégories. A cet effet, nous exploitons cette multitude pour diversifier les résultats retournés à l'utilisateur en prenant en considération la pertinence. Dans ce qui suit, nous allons présenter notre modèle de données et nous donnerons aussi les étapes de notre approche.

A. Modèle de donnée

Soit D l'ensemble de tous les documents du corpus, E l'ensemble des entités du corpus et K l'ensemble des mots clés. Chaque document $d \in D$ contient un ensemble d'entités $entites \in E$ et un ensemble de mots clés $\in K$. Nous supposons un ensemble de catégories C (Politique, Technology_Internet, Sports, ...) et un ensemble de types T (Person, Company, ...). Un document a un identificateur id , une catégorie $c \in C$ avec un score $score(c)$. Une entité a un identificateur id_e et a un ou plusieurs types $types \in T$ avec un score $score(e)$ du type.

B. Etapes de notre approche

Nous considérons D , corpus des documents semi ou non structurés (d'un forum, site de news, etc.), notre approche consiste à faire d'abord un prétraitement hors ligne pour préparer les informations au traitement en ligne, la figure (Fig. 4) résume les étapes de notre approche :

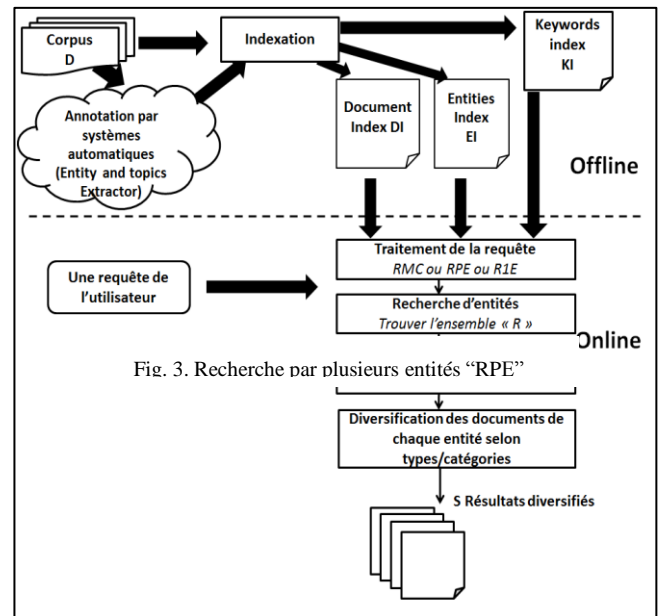


Fig. 3. Recherche par plusieurs entités "RPE"

Fig. 4. Etapes de notre approche

Etape 1 : Traitement Offline (Fig. 4)

Nous commençons par le traitement hors ligne qui consiste à annoter le corpus des documents D en utilisant un système d'annotation automatique tel qu'Open Calais¹⁰, pour extraire les entités, leurs types et les catégories des documents avec leurs scores d'extraction. Ce système rattache des méta-données sémantiquement riches aux documents en les catégorisant en thèmes et en extrayant les entités nommées qu'ils contiennent avec des scores.

Nous créons différents index pour stocker les informations, i.e. les mots clés, les entités, les types, les catégories et les scores.

Nous avons pris pour les scores : le $tf*idf$ pour le mot clé, le score(e) pour une entité e qui est extrait par Open Calais ainsi que le score(c) d'une catégorie extraite également par Open Calais. Trois index sont nécessaires, à savoir :

- Un index inversé classique pour les mots clés (KI : Keywords Index) qui stocke les mots clés avec leurs $tf*idf$.
- Un index inversé pour les types d'entités (EI : Entities Index), qui permet de déterminer les documents où une entité apparaît et quel est son type et son score (score(e))
- Un index pour les entités et les catégories des documents (DI : Document Index), qui permet de savoir quelles sont les entités d'un document et quelle est sa catégorie avec son score (score(c)).

Etape 2 : Traitement Online (Fig. 4)

Dans le traitement en ligne, nous utilisons nos index. Considérons une requête $Q = \{t_1, \dots, t_n\} / t_i \in E \cup K, Q$ est une requête de recherche par mots clés ou une requête de recherche par une/des entités. Le traitement des requêtes est présenté dans l'Algorithme 1.

Selon le type de la requête Q , notre idée est d'offrir une diversification d'interprétations en retournant un ensemble d'entités que nous nommons R , tel que :

¹⁰Open calais, <http://www.opencalais.com/calaisapi/>.

Cas 1 : R sont les entités qui apparaissent dans les documents communs entre les entités de la requête si la requête est de type RPE. Nous supposons que lorsque l'utilisateur exprime une requête d'entités (RPE), c'est qu'il cherche un lien ou veut faire une comparaison (ex, Sarkozy et Merkel, Renault ou Peugeot, infection et tumeur,...), voir l'Algorithme 1 suivant de la ligne 9 à la ligne 13.

Cas 2 : Un cas spécial de ce type de requête (RPE) est le cas Recherche par une Entité. Si la requête n'est composée que d'une seule entité, R sera égal à l'entité elle-même étendue par les entités composées par cette dernière, c'est à dire, les entités qui commencent, finissent ou contiennent l'entité de la requête. Voir l'Algorithme 1 de la ligne 4 à la ligne 7.

Notant que si la requête est un mélange d'entités et de mots clés (non vides et n'a pas d'opérateurs logiques), elle est alors considérée comme une requête de mots clés (RMC).

Cas 3 : R sont les entités qui apparaissent dans les meilleurs documents (Top k) qui répondent à la requête si la requête est une recherche par mots clés RMC. Les entités sont extraites en utilisant l'index DI (voir l'Algorithme 1 de la ligne 13 à la ligne 16). Pour ce type de recherche RMC, nous avons considéré la requête comme une requête Top k dont les résultats sont associés à un score. Dans notre travail le tf*idf du mot clé est utilisé comme score et est trouvé dans l'index KI. Les documents ayant le plus grand score sont les Top k résultats. Dans notre approche, nous avons utilisé l'algorithme du Top k processing proposé dans [18]. Nous supposons que les entités qui apparaissent dans les meilleurs documents ou les documents communs entre les entités de la requête sont pertinents (i.e. répondent à la requête) ou contextuelles (i.e. apparaissent dans le contexte donc peuvent intéresser l'utilisateur).

Après avoir trouvé l'ensemble R, un même traitement sera appliqué à toutes les entités trouvées pour diversifier les documents relatifs aux entités et pour que l'utilisateur puisse explorer les résultats de n'importe quelle entité de R.

Pour chaque entité $e \in R$, il s'agit d'identifier un ensemble $S = \{d_1 \dots d_m\}$ de documents à retourner à l'utilisateur comme suit :

– Si l'entité a plusieurs types : la diversification des documents est faite selon les types de l'entité. "Au moins un document" par type (meilleur document, selon le plus grand score(e)) doit être retourné à l'utilisateur pour garantir un maximum de diversité. Les autres documents seront sélectionnés selon pertinence c.-à-d., leurs scores doivent dépasser un seuil. Voir l'Algorithme 1 de la ligne 19 à la ligne 24.

– Si l'entité a un seul type : la diversification des documents est faite selon les catégories des documents relatifs à l'entité. "Au moins un document" par catégorie (meilleur document, selon le plus grand score) doit être retourné à l'utilisateur pour garantir un maximum de diversité. Les autres documents sont sélectionnés selon leur pertinence, c.-à-d., leurs scores doivent dépasser le seuil. Voir l'Algorithme 1 de la ligne 26 à la ligne 30.

La condition "Au moins un document" garantit que les documents ayant un type d'entité unique ou une catégorie unique ne seront pas négligés, même si leur score n'est pas élevé (n'atteint pas le seuil de pertinence). Cela maximise la diversité.

Dans notre travail, nous exploitons les différents types et catégories extraits par annotations du corpus et stockés préalablement dans des index. Ceci nous permet de contourner la complexité du problème de diversification des documents défini comme un problème NP-Complet puisqu'il s'agit de trouver un sous-ensemble divers de taille N dans un ensemble plus grand. C'est la définition de la diversité dans notre problème qui nous permet d'indexer le traitement et de simplifier la complexité des requêtes.

Algorithme 1 : Traitement_requête

Entrées : Q /*La requête*/

Sorties : R /*Ensemble d'entités*/,

S[] /*Les listes des documents diversifiés des entités*/

1 **Début**

2 R[] ; /*La liste des entités relatives trouvées*/

3 **Si** (Q.RPE() == Vrai) **alors**

4 **Si** (Q.nbr_entités() == 1) **alors**

5 /*Si l'utilisateur cherche par une seule entité*/ ;

6 R ← entités_composées(Q, EI) ;

7 /*Consulter l'index pour trouver les entités composées*/

8 **Sinon** /*Si l'utilisateur cherche par plusieurs entités*/

9 DocCommuns [] : liste ;

10 DocCommuns ← chercher_docs_communs (Q, EI)

11 R ← DI.entités(DocCommuns) ;

12 /*Trouver les entités qui apparaissent dans DocCommuns en utilisant DI*/

13 **Finsi**

14 **Sinon**

15 /*Requête de mots clés i.e. RPE == Faux*/

16 R ← DI.entités(topK(KI, Q));

17 **Finsi**

18 **Pour chaque e dans R do**

19 e.docs[] ← EI.Accès(e);

/*Mettre les documents de e dans docs */

20 **Si** (e.hasTypes == Vrai) **alors**

21 /*Si l'entité a plusieurs types */ ;

22 e.S ← types(meilleur_document (score(e))) ;

23 /*Prendre le meilleur document de chaque type de e pour assurer la diversité des types */

24 e.S ← e.S + documents_pertinents(seuil) ;

25 /*Ajouter les documents de chaque type qui dépassent le seuil de la pertinence */

26 **Sinon** /*Si l'entité a un seul type*/ ;

27 e.S ← catégories(meilleur_document (score(c))) ;

28 /*Prendre le meilleur document de chaque

catégorie pour assurer la diversité des catégories*/

29 e.S ← e.S + documents_pertinents (seuil) ;

/*Ajouter les documents de chaque catégorie qui dépassent le seuil de pertinence */

30 **Finsi**

31 **Fin**

32 **Fin**

V. EXPERIENCES

Nous avons implémenté en java un prototype du fonctionnement de notre système. L'architecture du système est composée de deux parties, une partie hors ligne (offline) et une partie en ligne (online). Le prototype réalisé permet dans un premier temps d'indexer et d'annoter le corpus de

documents afin de créer les index présentés préalablement (KI, EI, DI). Pour l'annotation, l'API d'Open Calais a été utilisée. Les index contiendront toutes les informations nécessaires au traitement en ligne des requêtes.

Dans le traitement en ligne, notre prototype permet de traiter les requêtes des utilisateurs. Il offre à l'utilisateur le choix entre une requête de recherche par entités ou une requête de recherche par mots clés et retourne des résultats, c'est à dire des entités et des documents pertinents aux entités trouvées. Les documents sont classés et diversifiés pour augmenter à la fois la pertinence et la variété des résultats.

Le corpus 20 NewsGroups¹¹ est utilisé pour les expériences. Ce dataset est une collection d'environ 20 000 messages, collectés de 20 différents newsgroups (environ 1000 messages par groupes). Ce corpus a été choisi pour sa richesse en entités et en catégories.

L'étude expérimentale est en cours, elle n'est pas totalement couverte par cet article, nous avons néanmoins mesuré la qualité des résultats obtenus et pour cela, nous avons fait des tests utilisateur. Par le biais de formulaires, nous avons demandé à 10 utilisateurs de numéroter de 1 à 3, les réponses de 5 requêtes obtenues par notre prototype, ((1) étant très pertinent, (2) peut être pertinent et (3) pas du tout pertinent), et ce pour les trois types de requêtes (R1E, RPE, RMC). Les réponses ont été agrégées et la moyenne a été calculée et transformée en pourcentage pour chaque requête (pourcentage de pertinence).

Requête R1E : Pour le premier type R1E, les requêtes posées sont présentées dans le tableau suivant ainsi que le nombre et les pourcentages des entités retrouvées :

REQUETES R1E

Requêtes R1E	Nombre d'entités trouvées	Pourcentage des entités pertinentes	Pourcentage des autres entités composées
Chevrolet	6	72.16%	16.66%
Lincoln	10	36.36%	48.48%
America	15	36.84%	42.10%
Ford	25	57%	35%
Cancer	74	31.03%	56.31%

Les entités composées sont pertinentes lorsque elles sont notées (1) par les utilisateurs, ceci représente le pourcentage des entités pertinentes (voir Table I). Nous avons remarqué que les entités sont parfois notées (2) puisque certaines entités composées sont moins pertinentes que les autres du point de vue de leur sens par rapport à l'utilisateur. Elles sont constituées par la même entité mais peuvent exprimer des choses complètement différentes (exemple : Lincoln : voiture, personne, ville) ce qui explique les résultats obtenus dans la Table I. Lorsque les entités trouvées sont notées (3), les utilisateurs nous ont exprimé le fait qu'ils ne connaissent pas l'entité composée retournée. Nous pensons enrichir les entités retournées par des snippets décrivant l'entité. Nous remarquons que le pourcentage moyen des requêtes de Table I est de 46.67% pour les entités pertinentes et 39.71% pour les autres entités

composées, donc globalement 87.38% des entités retournées satisfont les utilisateurs (très pertinentes ou peuvent être pertinentes).

Requête RPE : Pour le deuxième type RPE, nous présentons dans Table II les requêtes posées ainsi que le nombre et les pourcentages des entités trouvées :

REQUETES RPE

Requêtes RPE	Nombre d'entités relatives trouvées	Pourcentage des entités pertinentes	Pourcentage des entités contextuelles
Ford and Chevrolet	2	83.33%	16.66%
Lincoln and bush	3	66.66%	11%
Infection and tumors	8	50%	33%
Volvo USA	26	37.15%	25.61%
Washington and Baghdad	47	26.23%	27.65%

Lorsque les entités répondent à la requête elles ont été notées (1) (pourcentages de entités pertinentes, voir Table II), lorsqu'elles avaient un lien avec le contexte de la requête elles ont été notées (2) (pourcentages de entités contextuelles, voir Table II). Le reste des entités sont non significatives i.e. les entités qui apparaissent dans le contexte mais qui ne n'ont pas de liens avec la requête. Nous remarquons que le pourcentage moyen des requêtes de Table II est de 52.67% pour les entités pertinentes et 22.78% pour les entités contextuelles, donc globalement 75.54% des entités retournées satisfont les utilisateurs.

Requête RMC : Pour le troisième type RMC, les requêtes posées sont présentées dans Table III, nous présentons également le nombre des entités retrouvées ainsi que les pourcentages de pertinence :

REQUETES RMC

Requêtes RMC	Nombre d'entités relatives Trouvées	Pourcentage des entités pertinentes	Pourcentage des entités contextuelles
Ford car	10	45.21%	28.57%
Patient disease	11	39.02%	22%
Car dealer	12	31.47%	20%
Buy Ford	14	29%	29.32%
Prime minister	31	27.08%	35.82%

Dans Table III, nous avons calculé le pourcentage des entités pertinentes qui ont été notées (1) par les utilisateurs. Les entités notées (2) sont contextuelles (Voir Table III). Pour toutes les requêtes de Table III, au moins 27% des entités retournées sont pertinentes (requête prime minister) et au moins 20% des entités sont contextuelles (requête car dealer). Le pourcentage moyen des requêtes de Table III est de 34.35% pour les entités pertinentes et 27.14% pour les entités contextuelles, donc globalement 61.49% des entités retournées satisfont les utilisateurs. La qualité des résultats est moins bonne que les deux autres types de requêtes (R1E, RPE) car l'interprétation des mots clés est plus complexe que celle des entités. Une amélioration de cette partie de notre approche est prévue dans un futur travail.

¹¹<http://qwone.com/jason/20Newsgroups/>

VI. CONCLUSION ET PERSPECTIVES

Nous avons présenté une approche de diversification des résultats de la recherche des entités et nous avons considéré trois types de recherche : recherche par une entité (RIE), recherche par plusieurs entités (RPE) et recherche par mots clés (RMC). Le but de notre approche est de permettre l'exploration des résultats diversifiés qui sont retournés aux utilisateurs. Les entités pertinentes retournées (qui répondent à la requête) sont soit étendues par les entités composées si la requête est formée d'une seule entité soit par les entités apparaissant dans le contexte de la requête (entités contextuelles) si elle est formée par plusieurs entités ou par mots clés. Les entités trouvées peuvent avoir plusieurs types et les documents répondant à une entité trouvée peuvent aussi avoir plusieurs catégories. Notre définition de la diversité s'appuie donc sur les types et les catégories trouvés par annotation. Les documents les plus pertinents (dépassant un seuil) de chaque type ou catégories sont pris comme résultat pour assurer la diversité. En résumé nous pouvons dire que l'objectif de ce travail est d'introduire la notion de la diversité des résultats dans le cadre de la recherche d'entités et par entités, ainsi que la sélection des résultats les plus pertinents à différents types de requêtes. Pour une continuation de ce travail, plusieurs perspectives peuvent être envisagées, parmi ces perspectives nous pouvons citer :

- Attribution d'un nouveau score aux résultats en considérant la fiabilité de l'extraction des systèmes d'annotations.
- Améliorer les algorithmes pour éviter les cas d'impertinence des résultats.
- Appliquer un classement pertinent aux entités trouvées pour faciliter l'exploration et ajouter un descriptif aux entités retournées.

REFERENCES

- [1] T. Cheng, X. Yan, and K. C.-C. Chang. Supporting entity search: a largescale prototype search engine. In SIGMOD Conference, pages 1144-1146, 2007.
- [2] Agrawal R., Gollapudi S., Halverson A., Jeong S. (2009). Diversifying search results. In Wsdm, p. 5-14.
- [3] Drosou M., Pitoura E. (2010). Search result diversification. SIGMOD Record, vol. 39, no 1, p. 41-47.
- [4] Angel A., Koudas N. (2011). Efficient diversity-aware search. In Sigmod conference, p. 781-792.
- [5] T. Cheng and K. C.-C. Chang. Entity search engine: Towards agile best-effort information integration over the web. In CIDR 2007, Third Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 7-10, pages 108-113. www.crdldb.org, 2007.
- [6] H. Bast, A. Chitea, F. Suchanek, and I. Weber. Ester: efficient search on text, entities, and relations. In SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pages 671-678, New York, NY, USA, 2007. ACM
- [7] T. Cheng, X. Yan, and K. C.-C. Chang. Entityrank: searching entities directly and holistically. In VLDB '07: Proceedings of the 33rd international conference on Very large data bases, pages 387-398. VLDB Endowment, 2007.
- [8] Gianluca Demartini, [Claudiu S. Firan](#), [Mihai Georgescu](#), [Tereza Iofciu](#), [Ralf Krestel](#), [Wolfgang Nejdl](#): An Architecture for Finding Entities on the Web. [LA-WEB/CLIH 2009](#): 230-237.
- [9] M. Zhou, T. Cheng, and K. C.-C. Chang. Data-oriented content query system : Searching for data into text on the web. 2010.
- [10] T. Cheng and K. C.-C. Chang. Beyond pages : Supporting efficient, scalable entity search with dual-inversion index. 2010.
- [11] J. Peheveski, A.-M. Vercoustre, and J. A. Thom. Exploiting locality of wikipedia links in entity ranking. In C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, and R. W. White, editors, Advances in Information Retrieval , 30th European Conference on IR Research, ECIR 2008, Glasgow, UK, March 30-April 3, 2008. Proceedings, volume 4956 of Lecture Notes in Computer Science, pages 258-269. Springer, 2008.
- [12] G. Demartini, C. S. Firan, T. Iofciu, R. Krestel, and W. Nejdl. A model for ranking entities and its application to wikipedia. Web Congress, Latin American, 0:29-38, 2008.
- [13] G. Demartini, C. S. Firan, T. Iofciu, and W. Nejdl. Semantically enhanced entity ranking. In J. Bailey, D. Maier, K.-D. Schewe, B. Thalheim, and X. S. Wang, editors, WISE, volume 5175 of Lecture Notes in Computer Science, pages 176-188. Springer, 2008.
- [14] Bautin M., Skiena S. (2009). Concordance-based entity-oriented search. Web Intelligence and Agent Systems, vol. 7, no 4, p. 303-319.
- [15] Fafalios P., Kitsos I., Marketakis Y., Baldassarre C., Salampasis M., Tzitzikas Y. (2012). Websearching with entity mining at query time. In Irfc, p. 73-88.
- [16] Kitsos I., Magoutis K., Tzitzikas Y. (2013). Scalable entity-based summarization of web search results using mapreduce. In Distributed and parallel databases (dapd), springer journals.
- [17] Singhal A. (2012). Introducing the knowledge graph: things, not strings. <http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>
- [18] R. Fagin, A. Lotem, and M. Naor. Optimal aggregation algorithms for middleware. J. Comput. Syst. Sci., 66(4):614-656, 2003

Une approche sémantique pour la recommandation de services dans le cadre d'une formation en ligne

Khaled BEDJOU

Ecole Nationale Supérieure d'Informatique, ESI ex INI
Oued Smar, Alger, Algérie
K_bedjou@esi.dz

Lamia BERKANI

Département d'informatique
Laboratoire d'intelligence Artificielle (LRIA), USTHB,
Alger, Algérie
lberkani@usthb.dz

Faiçal AZOUAOU

Ecole Nationale Supérieure d'Informatique, ESI ex INI
Oued Smar, Alger, Algérie
F_azouaou@esi.dz

Résumé—L'utilisation des services web dans le domaine du e-Learning est devenue de plus en plus accrue. Cependant, vu le nombre important de ces services, une sélection adéquate est nécessaire pour améliorer leur apport et efficacité. Nous proposons dans cet article un système de recommandation personnalisée de services dans le cadre d'une formation en ligne. L'objectif est d'assister les apprenants dans leurs activités pédagogiques (exercices individuels, résolution de problèmes en groupe, etc.) en leur proposant les services les mieux adaptés, à la fois, à leurs profils et à leurs besoins. Le système est basé sur une approche sémantique en utilisant plusieurs ontologies. L'article présente la démarche globale de recommandation ainsi qu'une représentation sémantique des différentes ontologies utilisées. Afin d'illustrer notre travail, un prototype de système de recommandation a été présenté.

Mots clés—recommandation, service web, profil utilisateur, qualité de service, ontologie, e-learning.

I. INTRODUCTION

Les services web consistent à exposer sur un réseau, une ou plusieurs applications répondant à certains impératifs technologiques. Ces services peuvent proposer des fonctions très simples (du type requête/réponse) ou un ensemble complet d'outils, permettant d'aller jusqu'à la composition des services pour proposer une application complète.

Profitant de l'infrastructure du web sémantique, les services web dits « sémantiques » ouvrent des perspectives d'avancées clés dans le traitement d'information. Un large spectre d'applications peut être considéré, incluant des services dédiés au commerce en ligne (e-Enterprise, e-Business), à la recherche scientifique (e-Science), à l'éducation (e-Learning), ainsi que ceux dédiés aux citoyens (e-Government, e-Democracy), etc.

Le secteur de l'information et de la communication et plus particulièrement les plates-formes d'apprentissage à distance, ne sont pas restés indifférents à l'expansion des services web. Les exigences accrues des acteurs (Administrateurs, Enseignants, Apprenants) de ces systèmes en termes d'adaptation, de parcours, de partage et de réutilisation des contenus ont incité l'orientation de ces systèmes vers des environnements distribués basés sur des architectures orientées services web.

Une plate-forme d'apprentissage peut être vue comme un ensemble de web services qui coopèrent entre eux pour fournir certaines fonctionnalités aux acteurs de la

plateforme. Les architectures e-Learning orientées services web facilitent la réutilisation de fonctionnalités, l'interconnexion, la gestion et l'évolution d'applications à base de services.

Cependant, avec l'augmentation du nombre de services disponibles, il devient difficile de trouver les services les mieux adaptés à une situation d'apprentissage donnée. Pour cela nous nous intéressons dans le cadre de cet article à la recherche et recommandation personnalisées de services dans le cadre d'une formation en ligne.

L'objectif est de supporter les apprenants dans leurs activités pédagogiques en leur suggérant des services selon leurs profils et leurs besoins.

Nous proposons dans cet article une démarche globale de recommandation basée sur une approche sémantique. Un ensemble d'ontologies a été utilisées, dont : l'ontologie du profil utilisateur, l'ontologie de domaines de connaissances et l'ontologie dédiée aux services prenant en compte la dimension de qualité de services. Afin d'illustrer notre travail, nous avons proposé une architecture technique ainsi qu'un prototype de système de recommandation de services web pour une formation en ligne.

Le reste de cet article sera organisé comme suit : la section 2 décrit avec plus de détails la problématique de recherche. La section 3, présente un état de l'art sur les systèmes et approches de recommandation de services web, en particulier dans le domaine du e-learning. La section 4 propose notre solution de recommandation de services web dédiée aux apprenants dans le cadre d'une formation en ligne. La section 5 présente l'architecture du système ainsi que quelques captures écrans afin d'illustrer notre travail. Enfin, la conclusion, présente les apports de notre travail et présente des perspectives futures.

II. PROBLEMATIQUE

Une plate-forme pédagogique est un logiciel qui assiste la conduite des formations ouvertes et à distance. Elle est basée sur des techniques collaboratives et regroupe les outils nécessaires aux trois principaux acteurs de la formation : apprenant, tuteur, administrateur. Cependant, chaque utilisateur possède des besoins et des préférences différents. C'est pourquoi, il serait intéressant de personnaliser ces techniques, afin de fournir des résultats plus adaptés aux besoins des utilisateurs.

A partir de différents travaux, dont ceux présentés dans [1] et [2], nous identifions une liste non exhaustive

d'activités d'apprentissage dans un système e-Learning que nous avons classé selon l'acteur de l'activité en question (apprenant ou enseignant) comme suit:

- Les activités des enseignants (tuteurs) : regroupent la conception d'un contenu pédagogique, la publication et la mise à jour des contenus pédagogiques, les corrections des évaluations, etc.
- Les activités des apprenants : regroupent le suivi des cours, la réalisation des exercices, les tests d'évaluation, etc.

Ainsi, le fonctionnement est décomposé en différentes activités, ou groupes d'activités, qui peuvent être mises en application de façon indépendante sous la forme de services web.

Dans le cadre d'une formation en ligne, l'utilisateur est amené à utiliser des services web pour réaliser une tâche ou une activité d'apprentissage. Cependant avec la constante augmentation du nombre de ressources sur le web, l'accès à une ressource ou un service adapté aussi bien aux besoins de l'utilisateur qu'à ses préférences devient une tâche difficile. En effet, des outils de recherche délivrent des résultats où la quantité d'information est importante. L'utilisateur est donc, très vite submergé par des informations qui ne sont pas forcément en adéquation avec ses usages. C'est dans ce sens que la recommandation sémantique et personnalisée de services peut améliorer la qualité d'apprentissage en proposant à l'utilisateur les services web adéquats avec son contexte, ses préférences et ses besoins.

Notre objectif est de développer un système de recommandation de services web basée sur une approche sémantique. En effet, la représentation sémantique va nous permettre de mieux définir les besoins et profils des apprenants ainsi que le contexte de la formation. Pour cela, une étude des travaux déjà existants sera nécessaire.

III. ETAT DE L'ART

A. *Recommandation de services*

L'état de l'art montre que les approches de découvertes de services dépendent du niveau de représentation, sémantique ou syntaxique, des descriptions des services Web. Les approches de découverte de services web sont classées selon 3 grandes catégories d'approches [3] :

- **Des approches basées sur la syntaxe** : le principe général de cette approche est la comparaison syntaxique entre la requête, basée mots clés, de l'utilisateur et les descriptions syntaxiques (WSDL) des services Web. La découverte de service web dans cette catégorie d'approche est basée sur des techniques issues du domaine de la recherche d'informations (RI), tels que les travaux [4] et [5].
- **Des approches basées sur la sémantique** : se focalisent sur la description sémantique des services Web. Ce développement est de plus en plus significatif puisqu'il semble pouvoir aborder certaines insuffisances des approches basées sur les mots clés.

Les ontologies sont le modèle utilisé pour la représentation sémantique des services Web. Elles permettent d'établir des relations sémantiques entre les différents concepts d'un domaine. Parmi les travaux qui sont classés dans cette catégorie d'approche, nous citons les travaux de [6],[7],[8] et [9].

- **Des approches basées sur le contexte** : Le contexte d'un service Web peut grouper la localisation du service (restriction d'usage géographique du service), le coût d'utilisation, la catégorie de service, etc. Le contexte de la découverte de services est pris en compte afin de ne proposer à l'utilisateur que des services qui répondent au mieux à ses besoins et dont le contexte de chaque service offert est adéquat avec celui de l'utilisateur. Parmi les travaux qui sont classés dans cette catégorie d'approche, nous citons les travaux de [10], [11], [12], [13], [14] et [3].

B. *Recommandation de services en e-learning*

Dans le domaine de l'éducation ou formation en ligne, les travaux de découverte et recommandation de services sont plutôt classés dans les catégories d'approches sémantiques et contextuelles. Le nombre de plateformes e-learning, qui sont basées sur les web services, est de plus en plus croissant [15], [16] et [17].

Plusieurs travaux de recherches ont été réalisés sur la découverte de services web sémantiques dans le cadre de l'e-learning. Certains auteurs ont proposé une infrastructure pour la découverte de services basée sur l'utilisation d'une ontologie de e-learning. Ils ont exploité les relations sémantiques entre concepts et connaissances du domaine à enseigner pour permettre de retrouver les plates-formes d'e-learning disponibles. [18].

D'autres auteurs ont procédé à doter les services d'une description sémantique, à base d'ontologies, des différents critères de choix des services. Ainsi, d'élaborer une description ontologique du profil utilisateur, qui permet de filtrer les résultats de la recherche pour suggérer uniquement les résultats les plus pertinents. [19].

D'autres travaux se sont basés sur la logique floue et l'appariement multi-phases. L'objectif est de traiter l'information imprécise ou vague dans le service Web correspondant, et mettre en œuvre un raisonnement basé sur la logique floue et les variables linguistiques, et ainsi d'améliorer l'efficacité, le rappel et la précision de la découverte de services Web. [20].

IV. CONTRIBUTION

Nous proposons dans cette section, notre approche de recommandation sémantique de services dans le contexte d'une formation en ligne. Nous présentons tout d'abord le processus global de recommandation, puis nous exposons la représentation sémantique des différents aspects liés à notre recherche. Finalement, nous présentons l'architecture de notre système avec quelques scénarios d'utilisation.

A. *Processus de recommandation*

Suite à l'étude de l'état de l'art, nous avons constaté un manque de travaux sur la recommandation sémantique de

services. Pour cela, nous proposons une solution de recommandation de services, selon le processus suivant :

- Etape 0 : Authentification et récupération des informations du profil utilisateur.
- Etape 1 : renseigner les critères de recherche.
- Etape 2 : Recherche des descriptions des services web qui correspondent aux critères de la requête en utilisant l'ontologie des services web OWL-S qui sera décrite dans la section ci-dessous.
- Etape 3 : Pré-filtrage des résultats de recherche selon les critères de langues et préférences.
- Etape 4 : Filtrage des résultats trouvés selon les informations du profil utilisateur, grâce à l'ontologie du profil utilisateur.
- Etape 5 : Recommandation de service selon les informations de l'utilisateur et les services web et affichage des résultats de recommandation.

B. Représentation sémantique

Nous décrivons dans cette section la description des ontologies utilisées. Certaines de ces ontologies ont été développées en suivant la méthodologie de Noy [21], telles que l'ontologie du profil utilisateur et celle liée aux services. Tandis que d'autres sont déjà existantes et ont été exploitées dans notre recherche telle que l'ontologie de domaine de connaissances.

Nous décrivons ci-dessous ces ontologies comme suit :

1) Ontologie du profil utilisateur

Dans un domaine de l'apprentissage, toutes les dimensions de la personnalisation doivent être prises en compte pour répondre de manière pertinente et adaptée aux besoins d'un utilisateur donné (administrateur, enseignant ou apprenant). Nous citons en particulier les dimensions suivantes: données personnelles, centres d'intérêt, niveaux de connaissance, confidentialité des informations dans l'organisation de l'institution, contexte technique, contexte géographique, etc.

Afin de proposer une représentation sémantique du profil, nous nous sommes inspirés principalement de l'ontologie IMS-LIP [22] et de certaines ontologies présentées dans la littérature [19]. Les concepts pris en compte dans notre ontologie de profil sont les suivants :

- « *Info_personnelles* » : comporte les informations personnelles de l'utilisateur, à savoir son identité (nom, prénom, etc.), contacts (adresse, ville, email, etc.) et ses informations démographiques (date de naissance, langue maternelle, etc.).
- « *Education* » : spécifie les informations concernant le niveau d'éducation atteint par l'utilisateur. Elle englobe les domaines, le niveau et l'organisme d'éducation.
- « *Compétences* » : indique les domaines et les outils maîtrisés par l'utilisateur.
- « *Préférences* » : comporte les préférences des utilisateurs en ce qui concerne les langues, la licence (libre, payante), les régions où il souhaite que ses ressources soient disponibles, etc.
- « *Intérêts* » : indique les centres d'intérêts de l'utilisateur, soit dans le domaine de la formation ou dans d'autres domaines.

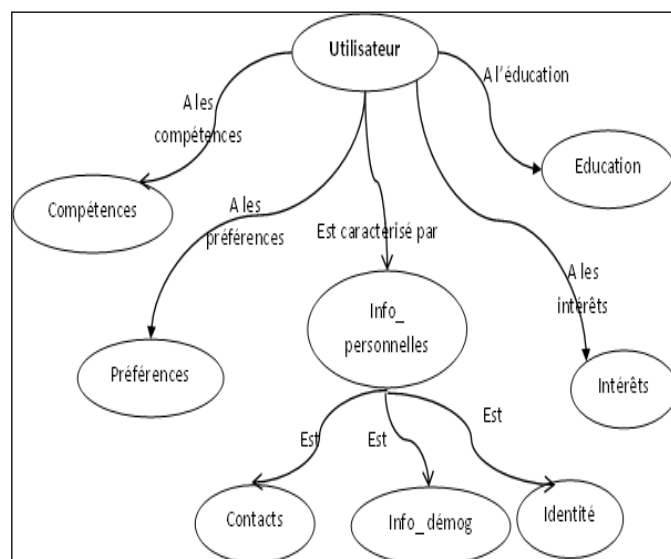


Fig. 2. Ontologie de profil utilisateur

2) Ontologie des services web

Le langage OWL-S (Web Ontology Language for Web services) est un sous-ensemble du langage OWL (Web Ontology Language) consacré à la description sémantique de services Web. Une description OWL-S est donnée en trois éléments [23] : le *service profile*, le *process model*, et le *grounding*. Ces trois éléments correspondent respectivement aux représentations suivantes : "que fait le service", "comment le service travaille" et "comment accéder au service".

La figure 2 ci-dessous représente une extension à cette ontologie pour pouvoir mesurer la qualité de service comme suit :

- « **Service profile** » : décrit les fonctionnalités des services Web. Il est utile pour leur découverte et leur sélection.
- « **Service model** » : détaille la sémantique des données échangées, au niveau des messages échangés entre services Web.
- « **Service grounding** » : indique l'encodage des données échangées, les protocoles de communication, ainsi que toutes les parties concrètes nécessaires à l'invocation du service.

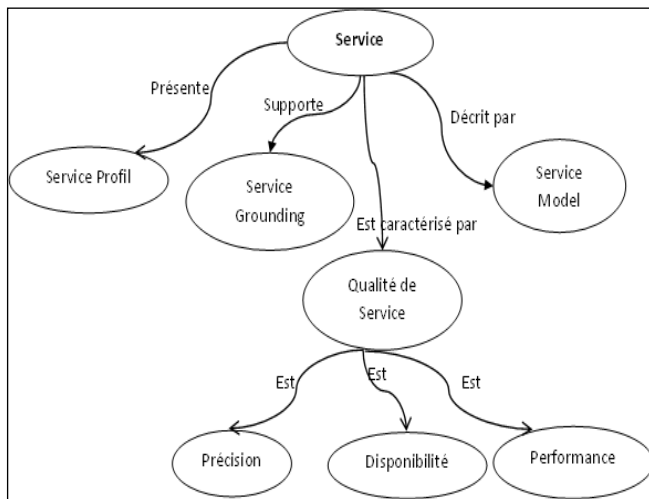


Fig. 3. Ontologie des services web

La qualité de service peut être déterminée selon ces paramètres :

- **Précision** : représente le degré d'exactitude de réponse d'un service à une requête donnée. Il est calculé par rapport au nombre de fois qu'il est invoqué, ainsi que le taux d'appréciation des utilisateurs. Par exemple, un service est invoqué 56 fois, et il y a 16 appréciations positives par rapport aux besoins des utilisateurs, dans ce cas, la précision sera égale à $16/56=0,28$.

- **Disponibilité** : représente la probabilité qu'un service soit disponible au moment de l'invoquer. Elle est mesurée par le pourcentage de temps qu'un service soit disponible sur une période prolongée. Par exemple, sur une période de 60 minutes (3600 secondes), un service X est invoqué 6 fois et le temps d'attente total du service était de 300 secondes. La disponibilité, dans ce cas, sera égale à $300/3600 = 0,083$.

- **Performance** : c'est la qualité de traitement des requêtes du service, mesurée par le nombre de requêtes traitées pendant une période et le temps nécessaire au traitement d'une requête. Par exemple, si le temps nécessaire de traitement de requête était de 25 secondes et que le nombre de requêtes traitées pendant 60 minutes (3600 secondes) était de 16 requêtes, la performance du service sera égale à $16*25/3600 = 0,111$.

3) Ontologie du domaine de connaissances

Nous utilisons également une ontologie de domaine de connaissances nommée DKOnto (Domain Knowledge Ontology), qui décrit les concepts liés aux domaines d'intérêts (exemple : génie logiciel, réseaux, bases de données, etc.). Nous considérons dans notre recherche l'ontologie hiérarchique du domaine de l'informatique qui dérive de la taxonomie connue d'ACM¹².

Étant donné que nous allons nous limiter dans cet article à la description du premier processus de recommandation (i.e. la recommandation directe), par conséquent, l'ontologie liée aux activités d'apprentissage sera présentée dans un prochain travail.

C. Application du processus de recommandation

La découverte de service, dans notre approche, est basée essentiellement sur les paramètres (Input, Output, Description du service Web, et Nom du service Web).

Une fois les termes de la requête de l'utilisateur sont récupérés, on procède au calcul de la similarité entre ces termes et les services web publiés dans l'ontologie OWL-S. Pour faire, nous avons deux cas de figures différents, le premier concerne la comparaison **terme-terme** (entre un **terme** de la requête d'un côté et les paramètres '**terme**' Input, Output, NameService de l'autre côté) ; le deuxième cas concerne la comparaison **terme-texte** (entre un **terme** de la requête d'un côté et le **texte** de la description du service web de l'autre côté). Nous avons proposé un algorithme de calcul pour chacun des deux cas comme suit :

1) Algorithme 1 :

Debut

T[] : tableau qui contient les termes de la requête

S[] : liste des services web publiés dans l'ontologie OWL-S

Pour chaque terme T[i] de la requête faire

Debut

Pour chaque service web S[j] de l'ontologie OWL-S

faire

Debut

P[] : paramètres du service web S[j]

Pour chaque paramètre P[k] faire

Debut

Result[] ← **CalculSimilarity1**(T[i], P[k])

FinPour

FinPour

FinPour

Fin

A la fin de l'exécution de cet algorithme, nous allons avoir comme résultats, un tableau Result[] de valeurs de similarité pour chaque terme de la requête. i.e. si nous avons n termes de la requête et m services web dans l'ontologie OWL-S, nous aurons n tableaux de 2 colonnes (termes/ Service) de m lignes chacun.

Pour la fonction **CalculSimilarity1** (T[i], P[k]), elle calcule le degré de similarité entre les deux termes T[i] et P[k], T[i] étant un terme de la requête et P[k] un des paramètres du service web.

Pour le choix de la mesure de similarité, nous avons opté au résultat de l'étude de comparaison des mesures de similarité sémantique les plus populaires sur l'ontologie WordNet, effectuée par [24], dans laquelle, la mesure de Jaing & Contrath [25] a donné les meilleurs résultats. Cette mesure combine entre les techniques basées sur les arcs et les techniques basées sur les nœuds. Cette mesure utilise la mesure de Resnik [26], la formule de calcul de similarité est donnée comme suit:

$$Sim_{jen}(X, Y) = \frac{1}{distance(X, Y)} \text{ with } distance(X, Y) = ic_{res}(X) + ic_{res}(Y) - 2Sim_{res}(X, Y) \quad (1)$$

En d'autres termes, la similarité est égale à l'inverse de la distance sémantique entre X et Y. La distance est égale au contenu en information de X + le contenu de Y – le contenu du parent le plus récent.

$ic_{res}(X) = -\log p(X)$: p(X) est la probabilité de trouver un concept ou un de ses descendants dans le corpus.

$ic_{res}(Y) = -\log p(Y)$: p(Y) est la probabilité de trouver un concept ou un de ses descendants dans le corpus.

¹²<http://www.acm.org/class/1998/>

$\text{Sim}_{\text{res}}(\mathbf{X}, \mathbf{Y}) = \text{Max} [\text{IC}(c)]$, c dans l'ensemble des antécédents des deux concepts X et Y où $\text{IC}(c)$ représente le concept le plus spécifique (qui maximise la valeur de similarité) qui subsume (situé à un niveau hiérarchique plus élevé) des deux concepts X et Y dans l'ontologie.

2) Algorithme 2 :

Nous présentons ci-dessous la description de notre second algorithme qui permet de calculer les mesures des similarités terme-texte. Dans ce cas, pour pouvoir calculer le degré de correspondance entre un terme de la requête et le texte de description de service web, nous allons utiliser les techniques de recherche d'information (IR) dans la fonction de calcul de similarité comme suit :

```

Debut
T[] : tableau qui contient les termes de la requête
S[] : liste des service web publiés dans l'ontologie OWL-S
Pour chaque terme T[i] de la requête faire
  Debut
  Pour chaque service web S[j] de l'ontologie OWL-S faire
  Debut
    text ← Description du service web S[j]
    Ponderation [] ← CalculSimilarity2 (T[i], text)
  FinPour
FinPour
TPonder [] ← (des poids des termes de la requête)
SPonder [] ← Ponderation []
Result ← Cosinus (TPonder[], SPonder[])

/* Cosinus (Rj, Sk) =  $\frac{\sum_i^N (W_{sij} * W_{rik})}{\sqrt{\sum_i W_{sij}^2 * \sum_i W_{rik}^2}}$  */
Fin

```

Pour la fonction **CalculSimilarity2** (T[i], text), elle calcule le degré de correspondance du terme T[i] au Texte text. Dans cette fonction la similarité est calculée en comparant la différence entre les descriptions textuelles de ces deux paramètres. La précision est améliorée en utilisant WordNet comme outil statistique donnant plus de poids à des termes très spécifiques étant positionnés profondément dans la hiérarchie. L'une des techniques utilisées dans le domaine de recherche d'information (RI) pour faire ce calcul est la norme TF-IDF.

TF-IDF compte la fréquence d'occurrence d'un terme dans un document par rapport à la fréquence d'apparition dans un corpus. Dans notre cas, on considère qu'un service dans l'ontologie OWL-S correspond à un document en RI, donc nous allons calculer un TF-s qui sera le nombre de fréquences d'un terme de la requête dans une description textuelle d'un service web. IDF est le nombre de fréquence d'un terme dans le corpus, dans notre cas, nous allons calculer IDF-s qui sera le nombre de fréquence d'un terme de la requête dans le corpus (l'ensemble des descriptions des services web)

Pour chaque terme de la requête, TF-s changera d'un service web à un autre, donc il sera calculé pour chaque service, tandis que IDF-s restera fixe pour l'ensemble des services, donc il ne sera calculé qu'une seule fois.

Pour un terme T_i de la requête, nous aurons

$$\text{IDF-}s_i = \log \left(1 + \frac{N}{N_i} \right) \quad (2)$$

Où N est le nombre total de services et N_i le nombre de descriptions de services dont T_i apparait.

$\text{TF-}s_{ij}$ = Nombre d'apparition du terme T_i dans le service S_j .

Les fréquences sont alors utilisées pour construire un vecteur pondéré de termes décrivant le service. Pour construire ce vecteur, nous calculons les pondérations globales pour chaque service comme suit :

$$W_{ij} = \text{TF-}s_{ij} * \text{IDF-}s_i \quad (3)$$

La similarité est donc déduite par le calcul de la similarité vectorielle entre les deux vecteurs de termes en utilisant une de ces techniques (Cosinus, Jaccard, Overlap, Dice, Euclidienne, etc.) issues du domaine de la Recherche d'Information.

En résumé des deux algorithmes que nous avons présentés, l'algorithme1 calcule le degré de correspondance des termes de la requête avec les paramètres (Input, Output et NameService) du service web, et rend comme résultat des tableaux de valeurs pour chaque terme. L'algorithme2 calcule le degré de correspondance des termes de la requête avec les descriptions des services web, et rend comme résultat des tableaux de valeurs pour chaque terme de la requête.

En faisant une réunion entre les tableaux par un calcul de moyenne pour chaque couple (Terme, Service), nous allons avoir un tableau global qui nous donne le degré de similarité entre la requête et les différents services web. Nous recommandons aux utilisateurs, les services dont le degré de similarité est supérieur ou égal à un seuil donné.

V. IMPLEMENTATION

Pour assurer le bon fonctionnement de notre système et mettre en œuvre les caractéristiques citées précédemment, nous proposons de développer une architecture basée sur un brassage entre le modèle J2EE (une plate-forme dédiée aux applications multi-tiers) et le modèle orienté services (client, fournisseur, UDDI). Nous avons opté pour ce choix parce que le J2EE est un langage complet, ouvert et il est le plus adapté et le plus utilisé avec les services web.

A. Architecture du système

Notre système est dédié aux consommateurs de services. Il assure la recherche et la recommandation des services qui répondent, le plus, à leurs besoins. Il présente deux types de recommandations : (1) Recommandation directe, en recherchant un service web ; et (2) Recommandation indirecte, durant la réalisation d'une activité d'apprentissage donnée.

Pour la recommandation directe, nous utilisons : les informations de l'utilisateur récupérées de l'ontologie du profil utilisateur, les descriptions des services web dans l'ontologie OWL-S et le module de découverte qui fait les calculs et matchings entre la requête et les services web.

Pour la recommandation indirecte, nous allons utiliser l'ontologie des activités d'apprentissage que nous allons décrire dans nos futurs travaux.

L'architecture technique du système est illustrée dans la figure 3 ci-dessous :

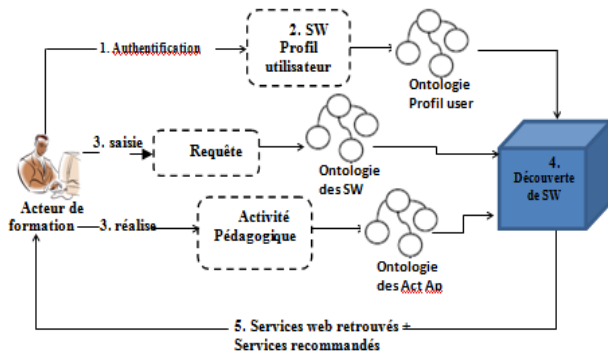


Fig. 3. Architecture technique du système

L'implémentation des ontologies que nous avons développées s'est effectuée à travers l'éditeur d'ontologies Protégé 3.5. Cet éditeur est open source, gratuit et possède une interface modulaire, ce qui permet son enrichissement par des modules additionnels (plugins). La figure 4 illustre un exemple d'une ontologie développée sous l'éditeur Protégé 3.5.

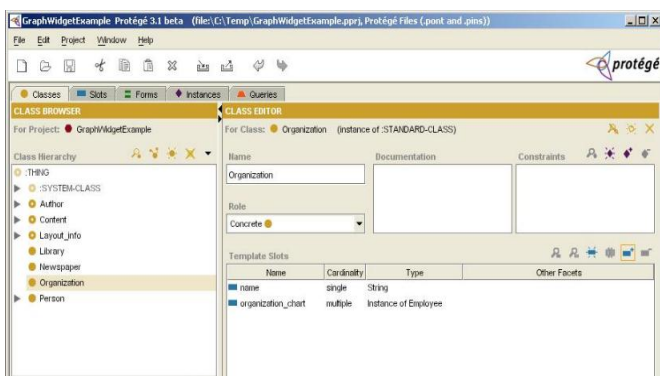


Fig. 4. Développement des ontologies sous Protégé

B. Prototype de système de recommandation de services web

Nous présentons dans cette section quelques captures de notre système de recommandation en les discutant selon le scénario suivant :

Nous supposons qu'un apprenant recherche des services web en introduisant un ensemble d'informations comme illustré par la figure 5.

Après authentification, le service web dédié à la gestion du profil utilisateur récupère les informations relatives au profil de l'apprenant : informations personnelles, préférences, centres d'intérêts...etc.

La recherche des services web se fait en introduisant la requête (mots clés) dans un seul champ, comme illustré dans la figure 5 ci-dessous :

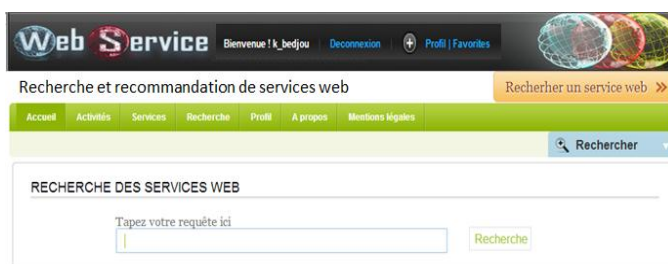


Fig.5. Page de recherche des services

Afin de pouvoir faire une recherche à base de mots clés saisis dans la requête, nous devons d'abord faire une analyse linguistique sur la requête. Deux méthodes du domaine de la recherche d'information sont proposées, la normalisation et la lemmatisation. Pour la normalisation il s'agit de garder les racines des mots (radical), quant à la lemmatisation, elle correspond à ramener les mots à leurs lemme (entrée au dictionnaire).

En général le radical des mots n'a pas de sens, et vu que dans notre cas, nous nous intéressons à la recherche sémantique (qui garde le sens des mots), nous avons opté à utiliser la méthode de la lemmatisation pour traiter les requêtes des utilisateurs.

L'extraction des lemmes est composée des étapes suivantes :

- Extraction des mots simples : suite de caractères séparés par (blanc, signe de ponctuation, caractères spéciaux, nombre).
- Suppression des mots vides (Stoplist, Common word removal) : mots trop fréquents mais pas utiles à la recherche.
- Détermination du lemme et de sa catégorie grammaticale pour chacun des mots à l'aide d'un lemmatiseur / analyseur syntaxique.
- Filtrage des lemmes en fonction de leur catégorie grammaticale. Seul les noms et les verbes sont conservés. Les autres lemmes sont supprimés.

Une fois la requête traitée, nous aurons une liste de termes (lemmes) avec lesquels nous procédons à la recherche sémantique des services web selon les deux algorithmes présentés précédemment, en faisant une comparaison (calcul de similarité) entre ces termes et les paramètres des services web.

Une fois les mesures de similarité calculées, le système peut retourner un nombre important de services web qui correspondent à la requête de l'utilisateur. Pour réduire le nombre de services retournés, nous procédons à un pré-filtrage des résultats selon les critères de langues et préférences de l'utilisateur c'est-à-dire ne pas recommander les services qui ne correspondent pas à la langue spécifiée par l'utilisateur et même chose pour les services qui peuvent ne pas l'intéresser en se référant à ces préférences.

Grâce à l'ontologie du profil utilisateur, nous allons rendre la découverte de service personnalisée i.e. si deux utilisateurs effectuent la même requête, le système propose des résultats différents pour chaque utilisateur et ce en utilisant l'ontologie du profil utilisateur pour filtrer les services web trouvés. Nous procédons donc, comme suit : le système accède à l'ontologie du profil utilisateur pour récupérer toutes ses informations (informations personnelles, intérêts, éducation, compétences...etc.) et avec ces informations nous filtrons les résultats trouvés dans l'étape précédente et ainsi nous proposons des résultats personnalisés pour chaque utilisateur.

Enfin, le système affiche les services recommandés pour l'utilisateur, comme illustré dans la figure 6 ci-dessous :

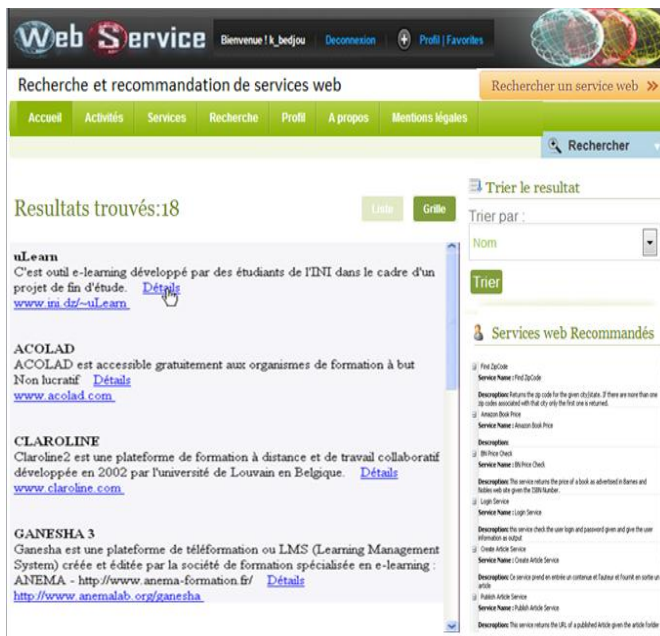


Fig. 6. Page de résultats de recherche

Notre système, affiche les résultats de recherches à gauche avec la possibilité de trier les résultats selon le choix de l'utilisateur (tri par nom de service, par catégorie de service gratuit ou payant, etc). La liste de services affichés à droite de la page représente des services suggérés par le système, ce sont des services qui peuvent intéresser l'utilisateur.

VI. CONCLUSION

Afin de faciliter l'accès aux services web dans le cadre d'une formation en ligne, nous nous sommes intéressés à une recommandation personnalisée basée sur une approche sémantique. Une recommandation sémantique de service web proposée à l'apprenant initialement ;

Nous avons proposé une démarche générale de recommandation. Comme première étape, nous nous sommes limités dans cet article à la recommandation directe, en recherchant un service donné, en présentant les différentes ontologies nécessaires. Le deuxième type de recommandation de services durant une activité d'apprentissage sera présenté dans nos prochains travaux.

De plus, nous envisageons de tester différentes fonctions de similarités sémantiques, et éventuellement de proposer notre propre métrique de similarité. De plus, il sera nécessaire de d'évaluer notre approche et de voir son intérêt pour les apprenants dans le cadre d'une formation réelle.

REFERENCES

- [1] Madjarov, I.: Des services web pour le e-Learning, Laboratoire des sciences de l'Information et des Systèmes, Maroc 2005
- [2] Vossen, G., Weterkamp, P.: E-Learning as Web Services 2003
- [3] Chelbabi, M.: Découverte de Services Web Sémantiques : une Approche basée sur le Contexte 2006
- [4] Newcomer, E.: Understanding Web Services- XML, WSDL, SOAP and UDDI, chapter 5, Finding Web Services : UDDI Registry. Addison Wesley Professional, May 2003
- [5] Romphong, P., Senivongse, T.: A query federation of uddi registries. In ISICT '03: Proceedings of the 1st international symposium on

Information and communication technologies, pages 561–566. Trinity College Dublin 2003

- [6] Paolucci, M., Kawamura, T., Payne, T.R., Sycara, K.P.: Semantic matching of web services capabilities. In ISWC '02 : Proceedings of the First International Semantic Web Conference on The Semantic Web, pages 333–347, London, UK 2002
- [7] Motta, E., Domingue, J., Cabral, L., Gaspari, M.: Irs-2 A framework and infrastructure for semantic web services. In International Semantic Web Conference, pages 306–318 2003
- [8] Vu, L.H., Hauswirth, M., Aberer, K.: Towards p2p-based semantic web service discovery with qos support. In Business Process Management Workshops, pages 18–31 2005
- [9] Verma, K., Mulye, R., Zhong, Z. Sivashanmugam, K., Sheth, A.: Speed-r Semantic p2p environment for diverse web service registries 2005
- [10] Strang, T., Popien, C-L.: A context modeling survey. In Workshop Proceedings, First International Workshop on Advanced Context Modelling, Reasoning And Management at UbiComp 2004
- [11] Pokraev, S., Koolwaaij, J., Wibbels, M.: Extending uddi with context-aware features based on semantic service descriptions. In ICWS, pages 184–190 2003
- [12] Keidl, M., Kemper, A.: Toward context-aware adaptable web services. In 13th World Wide Web Conference (WWW), New York, USA 2004
- [13] Doulkeridis, C., Loutas, N., Vazirgiannis, M.: A system architecture for context-aware service discovery. In International Workshop on Context for Web Services CWS-05 2005
- [14] Mostéfaoui, K.: Towards a contextualisation of service discovery and composition for pervasive environments. In Workshop on Web Services and Agent Based engineering 2003
- [15] Ivan, M.: Des services web pour le e-Learning. e-TI - la revue électronique des technologies d'information, Premier Numéro 2005
- [16] E-Learning as a Web Service. Seventh International Database Engineering and Applications Symposium (IDEAS'03) 2003
- [17] Xiaohong, Q., Anumit, J.: Web Service Architecture for e-Learning 2005
- [18] Addour, D.: Développement d'une plateforme pour la découverte de services web 2012
- [19] Boudali, F., Balla, A., Amrouche, H.: Découverte personnalisée des web services e-learning2008
- [20] Zhenglian, Haisong, Liang, Yonghua.: Framework of Semantic Web Service Discovery Based on Fuzzy Logic and Multi-phase Matching 2008
- [21] Natalya, F. Noy, Deborah, L., McGuinness.: Ontology Development 101: A Guide to Creating Your First Ontology 2000
- [22] MS Consortium, <http://www.imsproject.org>.
- [23] Martin, D.L., Paolucci, M., McIlraith, S.A., Burstein, M.H., McDermott, D.V., McGuinness, D.L., Parsia, B., Payne, T.R., Sabou, M., Solanki, M., Srinivasan, N., Sycara, K.P.: Bringing Semantics to Web Services :The OWL-S Approach. In J. Cardoso and A. P. Sheth, editors, SWSWPC, volume 3387 of Lecture Notes in Computer Science, pages 26– 42 2004
- [24] Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E., & Milios, E..Semantic similarity methods in wordNet and their application to information retrieval on the web.*Proc. of WIDM*, 10-16. 2005
- [25] Jiang, J., & Conrath, D. Semantic similarity based on corpus statistics and lexical taxonomy. *Proceedings of International Conference on Research in Computational Linguistics*. Taiwan 1997
- [26] Resnik, P. Semantic similarity in a taxonomy: an information based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11: 95-130. 1999

A Topology Management scheme with scalability and QoS guarantee for Large Scale Mobile Ad Hoc Networks in Urban Environment

Abdelhak Bentaleb¹, Saad Harous² and Abdelhak Boubetra¹

¹ Department of Computer Science, University of El Bachir El Ibrahimi, Bordj Bou Arreridj, Algeria

² College of Information Technology, United Arab Emirates University, P.O Box 15551, Al Ain, UAE
bentaleb_abdelhak@yahoo.com, harous@uaeu.ac.ae, boubetraabd@yahoo.fr

Abstract—in this paper we propose a scheme made up of two level hierarchies constructed based on trade-off between the clustering and virtual backbone techniques for large scale MANETs. We designed an efficient weight based clustering approach which considers QoS requirements in the first level. In the second level we built a virtual backbone based on connected dominating set algorithm. Our proposed scheme achieves several goals: support the network scalability when network size increases (thousands of nodes), supports the multimedia services with QoS, forms stable network, reduces the communications overhead, and improves the usage of scarce resources such as bandwidth.

Index Terms— Clustering; Large Scale MANETs; Weighted Metrics; Virtual Backbone; Dominating Set.

I. INTRODUCTION

Today, therecent rapid growth of cities and the evolution of wireless/mobile technologies [1], the world is becoming more and more urban [2]. The performance of the urban environment is affected by the availability of ICT-infrastructure and intelligent management [3][4]. Thus, Wireless and Mobile networks play an essential challenging role in the urban environments.

The mobile wireless networks are divided into two main classes. The most usual class named “cellular networks” that provides services by using a fixed infrastructures, typically consisting of a set of mobile nodes connected to each other. The second class, known as Mobile Ad Hoc NETWORKS (MANETs), which represents a self-configuring and self-organizing network of mobile devices connected by wireless links. Each device in a MANET is free to move independently in any direction, and will therefore change its links to other devices frequently without the support of centralized management [1]. Nearby nodes can communicate with each other directly by exploiting the wireless technologies of their devices in an ad hoc mode. Due to these characteristics, MANETs have gained great attraction recently and has been utilized to model problems in various fields and application domains such smart cities, military communications, forest hazards, hostile environments, disaster management, health applications, and emergency services.

Currently, with the emergence of multimedia services, network topology management with QoS, resource management and data routing have become hot research topics in MANETs that have attracted a lot of interest [1, 2]. However, it is very difficult to guarantee a good network topology management with QoS requirements of multimedia services in large scale MANETs. This is due to the

complexity of the features of these networks, namely: network size, limited bandwidth, dynamic topology, limited power, and lack of information about the network status, etc.

II. PROBLEM STATEMENT AND MOTIVATION

A. Problem Statement

In large scale MANETs, the network structuring (topology management) with scalability and QoS guarantee considered one of the important task to simplify the data routing operation of multimedia services. However, the topology changes unpredictably due to the nodes mobility [1] and also, maintaining the stability with low overheads in these kinds of networks is a challenging issue. Indeed, the development of feasible topology management schemes and protocols to manage large MANETs is a very complex task

The main issue in urban environments is how to support applications and services in a complex, distributed, large and diverse environment. This problem is being investigated intensively by many researchers [2, 3,5]. For some applications such as multimedia services (e.g., videoconference, VoIP, Video on Demand, etc.), most of the existing routing schemes are not suitable especially if there is a need to guarantee certain requirements of quality of service [13]. It is important to adapt MANETs to support network with large number of nodes and some acceptable level of QoS in order to efficiently deploy applications that are complex and require a lot of resources. To the best of our knowledge none of the proposed hierarchical routing structures [6, 7,9, 10,11, 14] support efficiently multimedia services in large scale MANETs.

B. Motivation and Goals

In last few years, several topology management schemes have been proposed focusing on different metrics [6, 14], to address the scalability issue in MANETs without taking into consideration the applications kinds and network size.

To address the aforementioned shortcomings, in this paper, we propose a network topology management scheme which satisfies the urban environment properties. Our proposed topology management scheme constructs two hierarchical levels based on the clustering and virtual backbone technique for large scale MANETs (Figure 1). The proposed scheme takes into account the network scalability with large number of nodes, and the QoS needed to support demanding applications such as multimedia services. In summary, our proposed scheme is a QoS-based hierarchical scheme for large scale MANETs in urban environment that is able to:

- Support the Scalability and QoS requirements of demanding applications like multimedia services.
- Maximize lifespan of the network.
- Minimize the number of clusters.
- Decrease the network overhead, create stable structure and improve the usage of scarce resources such as bandwidth.

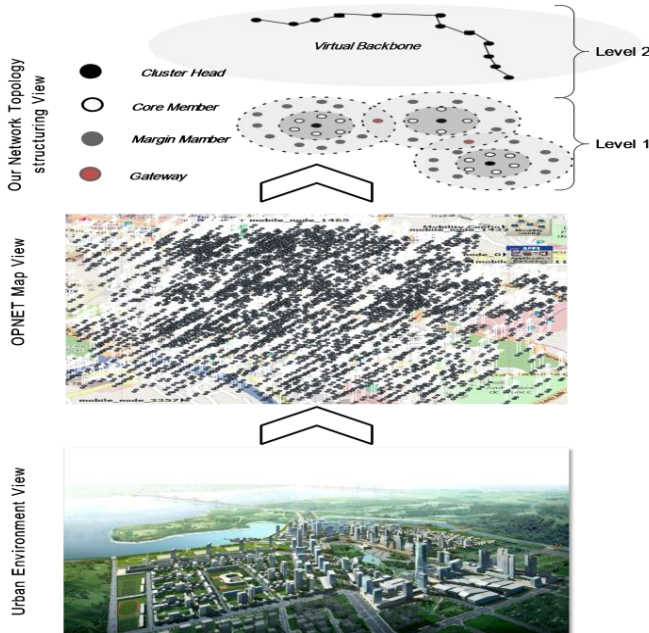


Figure 1: Our Network Topology Structure

The remainder of this paper is organized as follows. Section III reviews some related works, Section IV explains the proposed scheme and describes its phases. Section V analyses the performance of the proposed scheme by simulation. Finally, Section VI concludes the paper and introduces future work.

III. RELATED WORK

In the last few years, many research papers about network topology managements in MANETs have been published which aim to address the related issues in these kind of network such as: routing, QoS, scalability, security [6, 7, 8, 14]. Among those, the scalability with QoS guarantee of the multimedia services in large scale MANETs is one of the main challenges. To the best of our knowledge, most of the proposed schemes are based on the clustering technique or virtual backbone construction mechanism [6, 7, 10, 11, 12, 14, 20, 23, 25]. However, the majority of them resolve the scalability issue only without supporting the QoS of demanding applications like multimedia services. Also, they are not scalable when the network size increases to a very large number of nodes such as urban environment scenario. In other word, they do not take into consideration the applications kinds and network size.

In this section, we review some topology management schemes proposed in the literature.

In literature, various clustering schemes based and virtual backbone schemes based in MANETs have been proposed to organize the network, optimize the network communications, and ensure good information dissemination. A virtual structure is created to support the network services like routing, broadcasting, and optimize the usage of resources. Generally, these schemes differ on how they elect

cluster heads [6, 15]. Among these schemes we have Lowest-ID, Highest-Degree, Distributed Clustering Algorithm, Weighted Clustering Algorithm (WCA) and Distributed Weighted Clustering Algorithm (DWCA) schemes. Also on how the cluster heads are structured in the network [16]: Independent Dominating Set (IDS), Connected Dominating Set (CDS), Minimum Connected Dominating Set (MCDS) and Weakly Connected Dominating Set (WCDS).

In [17], Guizani et al proposed a new clustering algorithm in MANETs called α -Stability Structure Clustering (α -SSCA). α -SSCA is composed of three phases. The first phase consists in exchanging HELLO messages between neighbor nodes in order to collect the information necessary for cluster heads election. The second phase consists in CHs election and clusters formation. In this scheme, the election of cluster heads is based on the score metric, which represents the number of neighbors whose status has not been decided yet. Each node has one of the following statuses: CH, member node, not-decided. Initially all nodes are in the not-decided status. During cluster formation phase, the nodes having the highest score are declared as cluster heads and change their status to CH. After, each neighbor joins one of neighbor clusters as member nodes. Finally, the final phase is the maintenance of the cluster structure when there is a change due to nodes movement. This algorithm has some advantages like: it increase moderately the number of clusters with the aim of improving the topology stability of the clusters generated, reduces the overheads. The score metric is used in way to have the neighboring CHs far away from each other in order to eliminate cluster maintenance invocation when two cluster heads become one hop neighbors.

N. Mitton et al in [18] proposed two new ideas. First one, a new metric suitable to gather and aggregate network nodes into clusters and the second is a new distributed cluster head election heuristic called "Density". The proposed heuristic enables load balancing to guarantee a fair distribution of load among CHs, minimizing the number of control exchanged messages in the whole network. This heuristic is based on a criterion which gathers the neighborhood density of a node. This density criterion helps to keep the network stable when the topology evolves slightly (absorb small topology changes). The density metric takes into consideration both links and the number of nodes in a k -neighborhood. So, it is the ratio between the number of links and the number of nodes in a k -neighborhood. The k -density of a node u is defined as:

$$Pk(u) = \frac{|e = (v,w) \in E \vee v \in \{u, \Gamma k(u)\} \wedge w \in \Gamma k(u)|}{\delta k(u)}$$

Where: $Pk(u)$: the density of node u , $\Gamma k(u)$: list of k -neighborhood of a node u and $\delta k(u)$: degree of a node u (the number of k -neighbors).

During cluster formation, all nodes broadcast the HELLO messages periodically and build their k -neighbors lists based on the received HELLO messages. After, each node computes its k -density value using the density function and broadcasts it to all its k -neighbors in next HELLO message. The node with highest k -density value in its neighborhood is elected as a CH. In the case of a tie the algorithm uses the node ID as CH selection criteria. After CHs election process is done, each neighbor joins the cluster for which it has the highest density as member node. Each

member node broadcasts the cluster head ID and its density to its neighbor nodes. During cluster maintenance, each node periodically checks its density, neighbors and mobility. When node mobility or node density change events occur, the cluster maintenance phase will be invoked automatically and tries to adapt the network to all topology changes with less overhead. The proposed scheme generates a limited number of clusters and produces a less number of CHs re-election. However, this scheme does not support network with large number of nodes.

In [22] S. Guha and S. Khuller proposed two polynomial time approximation schemes for minimum connected dominating set (MCDS) problem which they based on finding a connected dominating set of minimum size. These schemes are centralized-based. The first scheme is characterized by a greedy algorithm which has the approximation ratio of $2(H(\Delta) + 1)$ where Δ is the maximum degree and H is the harmonic function. This scheme builds a spanning tree T based on four main steps. The node with the highest degree is selected as the root T . Initially, all nodes are marked with *white*. After, the node with the largest number of *white* neighbors is marked as *black* and its neighbors are marked as *grey*. The *black* and *grey* nodes then become members of T . The algorithm repeats until no *white* node exists in the network. The non-leaf nodes of T form the CDS (the black nodes). The second scheme is an enhancement of the first scheme. It is divided into two phases. During the first phase, the algorithm finds a dominating set. During the second phase it connects the dominating set using a Steiner tree algorithm [19]. The approximation factor of this algorithm is $H(\Delta) + 2$ with a ratio of $3 + \ln(\Delta)$.

J. Wu and H. Li [21] proposed a distributed pruning-based algorithm for calculating connected dominating set in MANETs. The proposed scheme is based on the two hops neighbors information knowledge. This algorithm uses a marking process where each node marks itself as a CDS member if it has two unconnected neighbors. This algorithm creates a virtual backbone but adds many redundant nodes. The authors presented an enhancement of this algorithm by adding two dominant pruning rules named marking rules. These rules reduce the size of the CDS by removing some redundant CDS members.

In [22], F. Dai et al proposed an extended and generalization version of [21] in order to reduce the size of the dominating set in the network. The proposed dominant pruning rule referred to as *Rule k* which extended the pruning rules [21] to *k-hop* neighborhoods in order to achieve better results.

IV. OUR PROPOSED APPROACH

Our scheme builds two hierarchical levels based on a trade-off between the clustering approach in the first level and the virtual backbone techniques in the second level for large scale MANETs.

Node States

In our scheme, we define five states, each node is in one of the following states: Cluster head (CH), core member (COM), margin member (CAM), not-decided (ND), gateway node (GW) depending on its roles (Figure 2). Initially, all nodes in ND status.

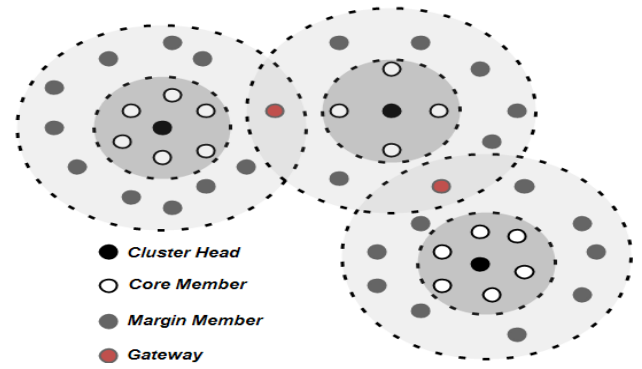


Figure 2: General Structure and node states in our Network Topology

Cluster Formation Phase

Our cluster formation phase consists of four main stages to build the clusters structure:

1. Neighbor Nodes detection

Initially all network nodes are in *not-decided* state. During this stage, all network nodes exchange a D-HELLO message periodically so that each node notifies its neighbor nodes of its presence and builds/maintains its local topology information.

2. Cluster heads election process

After neighbor nodes detection stage the cluster head election process is invoked. This stage is dedicated to the cluster heads election in the whole network. In the following, we propose a cluster head election algorithm that allows selecting a set of robust and optimal cluster heads in large scale network and dividing the network into a number of clusters. During this stage, each node v_i computes its weight value W_i and broadcasts it to its k -neighbors in the next *D-HELLO* message. Then, on reception of *D-HELLO* message, each node v_i compares the received weight values of its neighbors with its own weight value. If its own weight value is the highest, it declares itself a cluster head. Otherwise, it chooses the neighbor having the highest weight value among all its k -neighbors as its cluster head. The node that has many votes (selected by several neighbors as CH in its neighborhood), chosen as a cluster head in its neighborhood. Each node v_i based on a combination of QoS metrics and clustering metrics to compute its weight value W .

Clustering metrics:

Node trust (T): the trust value T_{v_i} of node v_i represents the direct relations relating to its 1-hop neighbors and indirect relations relating to its k -hop neighbors. In our algorithm we used the method described in [28] to calculate the trust node value (T).

Node density (D): the node density D_{v_i} of node v_i represents the ratio between the number of links and the number of nodes in a k -neighborhood. We used the model described in [18] to calculate the density metric (D).

Node mobility (M): In our algorithm, each node v_i uses the method described in [29] to compute its mobility value (M_{v_i}). The mobility value (M_{v_i}) of node v_i represents its effective average relative speed.

Battery remaining energy (E): The energy value E_{v_i} of node v_i is a serious metric that affects directly the lifetime of the cluster structure and network which represents the battery remaining power of node v_i .

After the calculation of four metrics (T, D, M, E), the clustering weight value $Cl(v_i)$ of a node v_i is calculated as:

$$Cl(v_i) = \omega_1 T_{v_i} + \omega_2 D_{v_i} + \omega_3 M_{v_i} + \omega_4 E_{v_i}$$

Where ω_i are weight factors for the corresponding metrics with $\sum \omega_i = 1$

QoS metrics:

Local Available Bandwidth (LBW): The local available bandwidth LBW_{v_i} of node v_i represents the unconsumed bandwidth at node v_i . Each node v_i defines its own LBW_{v_i} by passively listening to network activities. In our algorithm, we use the model described in [26] to calculate the Local Available Bandwidth (LBW).

Link Quality (S): The link quality value S_{v_i} of node v_i is defined as the estimated number of transmissions required to successfully send a packet over the link. We used the model described in [30] to calculate the link quality value (S).

After the calculation of four metrics (T, D, M, E), the QoS weight value $QoS(v_i)$ of a node v_i is calculated as:

$$QoS(v_i) = \omega'_1 LBW_{v_i} + \omega'_2 S_{v_i} + \omega'_3 \text{node ID}$$

Where ω'_i are weight factors for the corresponding metrics with $\sum \omega'_i = 1$

After that, each node v_i computes its weight value based on a combination of QoS metrics and clustering metrics and selects the optimal cluster head that has the highest weight value W in its neighborhood.

$$W(v_i) = w_1 Cl(v_i) + w_2 QoS(v_i).$$

Where w_i are weight factors for the corresponding metrics with $\sum w_i = 1$.

3. Network node joins process:

During this stage, each node joins a suitable cluster. Node v_j joins the neighbor cluster head that has the largest weight value in its neighborhood.

4. Gateway election process:

Once cluster heads are elected and nodes join process is completed, each cluster head selects its gateway nodes to communicate with neighbor clusters based on the QoS metrics. Nodes are candidate to be gateways if they have margin member status and they hear two or more cluster heads. The gateway node g with highest QoS value ($QoS(g)$) among other gateway nodes candidates is selected.

B.3. Virtual Backbone Construction phase

Once the gateways are selected in the network and the clusters are formed, the virtual backbone construction algorithm is invoked in order to build a virtual backbone in the network at level two. Our virtual backbone construction algorithm is divided into two parts. The first one is a distributed algorithm to find the optimal dominating set (DS) which covers all member nodes in the whole network. The second part is a connectivity algorithm to connect the dominating set and form the Connected Dominating Set. A virtual backbone can be formed by constructing a Connected Dominating Set (CDS).

1. Dominating Set Construction and connectivity:

Once the first level is done and the clusters structure is generated, the connected dominating set step is invoked in order to select the set of CDS and generate the virtual backbone. Initially, all cluster heads and gateways nodes are selected as DS member (part 1). After, each cluster head selects a set of connectors from its member nodes based on two metrics: the QoS value (the one with high QoS value from its member nodes) and (k, m) rule in order to ensure the connectivity between DS members (part 2). This mechanism leads to construct a partial virtual backbone in each cluster (optimal dominating set (DS) which covers all member nodes in each cluster). After that, the set of partial VB are connecting using the clusters gateways to construct the complete VB.

B.4. Topology Maintenance

Our topology maintenance tries to adapt our network structure to all topology changes that can occur due to nodes mobility. We define several types of events for topology maintenance invocation. In the first level we define three types of event: The node movements, the trust threshold property and the cluster head weight value change.

In the second level, some DS nodes may cause link failures due to dynamic topology that lead to the virtual backbone structure imbalance and network instability. Therefore, we introduce a new mechanism named dominating set recovery mechanism which is capable to select alternative dominating nodes from the set of DS nodes candidate.

Our proposed topology maintenance mechanism aims to keep the network connected and reduce the number of re-elections. Thus, we are able to maintain the network stability during the topology construction for large scale MANETs.

V. PERFORMANCE EVALUATION

In this section, we present the initial performance evaluation of our proposed scheme using Optimum Network Performance simulator (OPNET) [27]. In Table 1, we provide all simulation parameters.

A. Simulation model and parameters

The network model that we designed to simulate and evaluate our scheme consists of 500 - 3000 mobile nodes of type MANET placed randomly in an area of simulation 2000×2000 meters square. The mobility model we have chosen is the model RWP (Random Way Point), node speed between 0 to 100 m/s, IEEE 802.11 as Physical and MAC layer.

We assume that the radio model uses data-rate of 2 Mbits/s, Transmission range 250 meters, packet size 128 bytes, simulation time is 900 seconds.

Table 1. Simulation Parameters

Parameter	Meaning	Value
N	Number of Nodes	[500 – 3000] nodes
X × Y	Area of simulation	2000 m × 2000 m
Speed	Speed of the nodes	0 m/s - 100 m/s
Mobility	Mobility Model	RWP
Tx	Transmission range	250 m
PT	Pause Time	0 sec
HI	Hello Interval	2 sec
Pkt Size	Packet Size	128 byte
Data rate	Data rate	2 Mbits/s
W _i	Weights	Equally
Duration	Simulation Time	900 sec

In our initial experimental study we validate the scalability and QoS of our proposed scheme in very large network in terms of number of generated clusters and average lifetime duration of cluster, average number of CH changes, total overheads, clusters stability, and number of dominating set.

The main goal of our scheme is to support the network scalability when network size increases (thousands of nodes), supports the multimedia services with an acceptable level of QoS. So, in this part, we present the initial simulation results (Figure 3) of our network topology management scheme, and its comparison with previous schemes [6, 17, 18, 22, 24]. Based on the initial simulation results, our scheme performed better than others schemes.

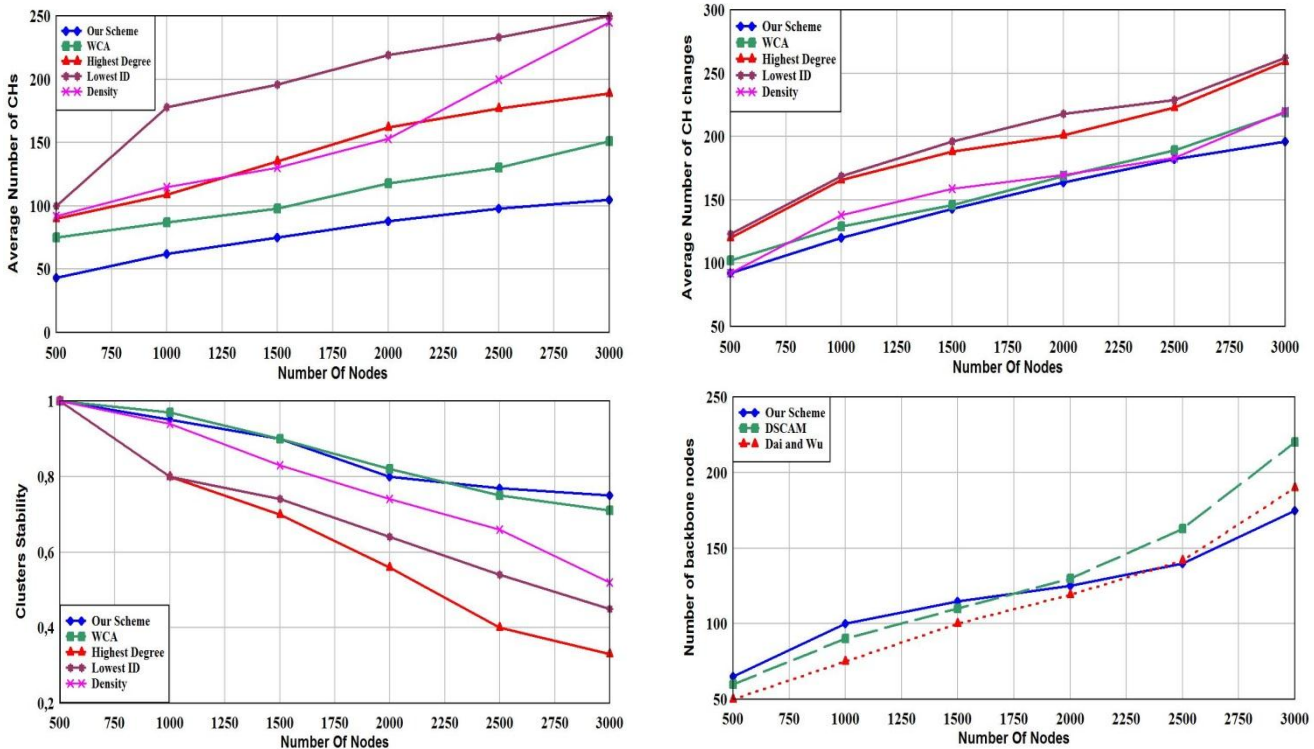


Figure 3: The Initial simulation results

VI. CONCLUSION

In this work, we studied the scalability and QoS issues for large scale mobile ad hoc networks in urban environment. Network topology management approach is one of the solutions that got a great attention in last few years. In this paper, we proposed a QoS based hierarchical scheme for large scale MANETs within urban environment scenario. Our scheme considers a trade-off between clustering approach and virtual backbone approach in order to build a 2-level hierarchical scheme.

We compared the performance of our scheme with a number of existing network topology management schemes [6, 17, 18, 22, 24]. The preliminary simulation results (Figure 3) have shown that our scheme performs generally better than the other network topology schemes in terms of: average number of CHs, average number of CH changes, total overheads, and number of backbone nodes. As future work, we plan to run simulation with very large number of nodes (up to 3000 nodes) and study the effect of other metrics.

REFERENCES

- [1] Corson, S., Macker J. (1999), "Mobile Ad hoc Networking", <http://tools.ietf.org/html/rfc2501>.
- [2] Augé-Blum, I., Boussetta, K., Rivano, H. (2012), "Capillary Networks: A Novel Networking Paradigm for Urban Environments", the first workshop on Urban networking, pp.25-30.
- [3] Caragliu, A., Del Bo, C., and Nijkamp, P., (2009). "Smart cities in Europe", Series Research Memoranda 0048.
- [4] European Smart Cities project, <http://www.smart-cities.eu>.
- [5] Woo LEE, Y., (2013), "Ubiquitous (Smart) City", EU Parliament Seminar.
- [6] Bentaleb, A., Boubetra, A., Harous, S. (2013), "Survey of Clustering Schemes in Mobile Ad hoc Networks", Communications and Network, Vol. 5 No. 2B, pp. 8-14.
- [7] Correa, B.A., Hincapie, R.C., Ospina, L. (2007), "Survey on Clustering Techniques for Mobile Ad Hoc Networks", Rev. Fac. Ing. Univ. Antioquia, No. 41, pp. 145-161.
- [8] Suman, P., Bisen, D., Tomar, P., Sejwar, V., Shukla, R. (2009), "Comparative study of Routing Protocols for Mobile Ad-Hoc Networks", International Journal of Information Technology & Knowledge Management, Vol. 4 No. 1, pp. 40-46.
- [9] Hong, X., Xu, K., and Gerla, M. (2002), "Scalable Routing Protocols for Mobile Ad Hoc Networks", IEEE Network magazine, Vol. 16 No. 4, pp. 11-21.
- [10] Butenko, S., Cheng, X., Zhu, D., and Pardalos, P. (2003) "On the construction of virtual backbone for ad hoc wireless network". Book chapter: Cooperative Control: Models, Applications and Algorithms Vol. 1, pp. 43-54.
- [11] Basagni, S., Mastrogiovanni, M., Panconesi, A., and Petrioli, C. (2006) "Localized protocols for ad hoc clustering and backbone formation: a performance comparison. Parallel and Distributed Systems", IEEE Transactions, Vol 17 No 4, pp.292-306.
- [12] Lee, B., Yu, C., Moh, S. (2005), "Issues in scalable clustered network architecture for mobile ad Hoc networks", Handbook of Mobile Computing, CRC Press, pp.611-637.
- [13] Crawley, E., Nair, R., Rajagopalan, B., Sandick; H., (1998) "A Framework for QoS based Routing in the Internet": <ftp://ftp.nordu.net/rfc/rfc2386.txt> (RFC 2386).
- [14] Bao, L., Garcia-Luna-Aceves, J. (2003) "Topology management in ad hoc networks", 4th ACM international symposium on Mobile ad hoc networking & computing, pp.129-140.
- [15] Bentaleb, A., Harous, S. and Boubetra, A. (2013), "A Weight Based Clustering Scheme for Mobile Ad hoc Networks", (MoMM2013), Vienna, Austria, pp.161-167.
- [16] Cokuslu, D., Erciyas, K., Dagdeviren, O. (2006), "A Dominating Set Based Clustering Algorithm for Mobile Ad Hoc Networks". Book chapter: Computational Science, Vol. 3991, pp. 571-578.
- [17] Guizani, B., Ayeb, B., Koukam, A. (2011), "Impact of Stability in Cluster Based Link State Routing Protocol for Self-Organizing Networks," 7th ICWMC.
- [18] Mitton, N., Busson, A., Fleury, E. (2004), "Self-organization in large scale ad Hoc networks", In Med-Hoc-Net, Brodum, Turkey, pp. 14-23.
- [19] Yu, J., Wang, N., Wang, G., Yu, D. (2013), "Connected dominating sets in wireless ad hoc and sensor networks: A comprehensive survey". Computer Communications, Vol. 36 No. 2, pp. 121-134.
- [20] Guha, S., Khuller, S., (1998), "Approximation Algorithms for Connected Dominating Sets", Algorithmica, Vol.20 No.4, pp.374-387.
- [21] Wu, J., Li, H., (1999), "On Calculating Connected Dominating Sets for Efficient Routing in Ad Hoc Wireless Networks"; 3rd international workshop on Discrete algorithms and methods for mobile computing and communication, pp. 7-14.
- [22] Dai, F., Wu, J., (2004) "An extended localized algorithm for connected dominating set formation in ad hoc wireless networks", IEEE Transaction on Parallel Distributed Systems Vol.15 No.10, pp.908-920.
- [23] Anitha, V.S., Sebastian, M.P., (2009), "SCAM: Scenario-based Clustering Algorithm for Mobile ad hoc networks", First International Communication Systems and Networks and Workshops, pp. 1-8.
- [24] Anitha, V.S., Sebastian, M.P., (2009), "Scenario-Based Diameter-Bounded Algorithm for Cluster Creation and Management in Mobile Ad hoc Networks", 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications, pp. 97-104.
- [25] Bentaleb, A., Harous, S., Boubetra, A. (2014), "A scalable clustering scheme and its performance evaluation", International Journal of Pervasive Computing and Communications, Vol. 10 No. 1, pp.27-42.
- [26] Yang, Y., Kravets, R., (2005), "Contention-aware admission control for ad hoc networks". IEEE Transactions on Mobile Computing, Vol. 4 No. 4, pp.363 - 377.
- [27] OPNET technologies, inc. <http://www.opnet.com>. 2009.
- [28] Babu, S., Raha, A., Naskar, K. (2011) "Geometric Mean based Trust Management System for WSNs (GMTMS)", In IEEE Proceedings of World Congress on Information and Communication Technologies (WICT 11), Mumbai, India, pp. 444-449.
- [29] Ni, M., Zhong, Z., Zhao, D. (2011), "MPBC: A Mobility Prediction-Based Clustering Scheme for Ad Hoc Networks", IEEE Transactions on Vehicular Technology, Vol. 60 No. 9, pp: 4549- 4559.
- [30] De Couto, D., Aguayo, D., Bicket, J., Morris, R. (2003) "A high-throughput path metric for multi-hop wireless routing," 9th Annual International Conference on Mobile Computing and Networking, pp: 134-146.

Overview of Incentive Mechanisms in Peer-to-Peer Systems

Fatima Lamia Haddi and Mahfoud Benchaïba

Laboratory of Information Systems (LSI), Department of Computer Science
University of Science and Technology Houari Boumédiène
Algiers, Algeria
{flhaddi, mbenchaiba}@usthb.dz

Abstract—Peer-to-Peer (P2P) systems suffer from non-contribution of peers. Non-contribution has several forms depending on the P2P application and system types and hence peer's possible collaborative operations. Peers are usually driven by their self-interest at the expense of social welfare. The main objective of an incentive mechanism is to motivate peers to collaborate to simultaneously achieve self-utility and social welfare. Multiple incentive mechanisms have been proposed to enhance collaboration. In this paper, we give a detailed definition of an incentive mechanism's objectives and roles. We then present a developed classification of incentive mechanisms in P2P systems. We also survey some practical incentive mechanisms in P2P systems and finally discuss their implementation.

Keywords—Incentive mechanisms; P2P systems; freeriding; collaboration enhancement.

I. INTRODUCTION

In a Peer-to-Peer(P2P) system, a peer can ask and offer services to other peers. Peers connect and self-organize in an overlay network and resources are deployed all over the system. Offering logical (application oriented) services involves physical resources (CPU, memory and bandwidth).

In designing P2P systems, a common assumption is an altruistic behavior of peers. All peers are supposed to unconditionally contribute resources and respect protocols; however, they are classified according to their behaviors as *contributors* and *non-contributors* since they are either motivated by altruism, rationality, selfishness or malice. Moreover, their motivation and behavior can evolve with time. Peer's motivations for non-contribution are resource saving for personal using or malice. Malicious behavior appears when peers become competitors or in order to collapse the entire system.

Contribution implies costs peer wise and brings benefits peer and system wise. Non-contribution is encouraged by decentralization, autonomy and hidden motivations (the fact to hide the reasons of defection to the other peers). Peers are driven by users, who often only consider their own interests and utilities, at the expense of the system's welfare and performance [1]. Low contribution implies a global resources' scarcity, a reduced system's attractiveness, less robustness, less scalability, less fault tolerance, important failure rate and may lead to the *tragedy of commons*: the peers demanding resources are competing against each other to access them while those offering the resources are overwhelmed and cannot respond. Non-contribution has no direct consequences on a peer's rights: it can still ask for resources and consumes them. The peer wishes to maximize its benefits and minimize its participation costs. Systems

might tolerate a certain non-contribution level but collapse if no peer shares any resource. Contribution usually incurs a cost at the provider peer's level: giving up some of its physical resources for a period and not using them for own purposes. Since system's viability and performance are linked to collaboration, an extrinsic motivation is to be created to stimulate peers to provide at least as much resources they consume, reducing the burden on the system in terms of resources consumption.

Our contributions, in comparison with other papers as [1] are:

- a clear definition of an incentive mechanism's objectives and roles,
- a classification of incentive mechanisms according to the targeted application or system and an in-depth definition of the incentive schemes,
- a highlight of the desirable features of an incentive mechanism.

This paper's outline is as follows. We first state the roles and the objectives of incentive mechanisms in section 2. We then introduce a developed classification of incentive mechanisms based on the observed operations of the system or the application they target, review some proposed incentive schemes, discuss their advantages and drawbacks and present some examples proposed in the literature in section 3. In section 4, we underline the main points of the proposition of a practical incentive mechanism. Finally, we conclude in section 5.

II. OBJECTIVES AND ROLES OF AN INCENTIVE MECHANISM

A. Incentive mechanisms objectives

The objectives of an incentive mechanism are the following:

- ensuring system's viability,
- making contribution beneficial to peers,
- balancing resource consumption and provision at a peer,
- reducing or eliminating non-contribution possibility,
- ensuring fairness between peers in getting compensation,
- ensuring resource abundance.

A P2P system's viability inherently depends on resource sharing. Altruistic behavior is desirable to system's operability in the introduction phase. The incentive

mechanism might guarantee that every newcomer contributes and should make contribution beneficial.

An incentive mechanism's major aim is to reduce or eliminate non-contribution. Nevertheless, an equal peers' contribution is not necessary [2]. The system might tolerate a certain amount of defection below a determined threshold, under which system's performance is not majorly affected (according to the required QoS in the system) and might decide to give defectors a redemption chance.

An incentive mechanism should ensure fairness: benefits are balanced between peers according to their contribution levels when peers with different capabilities are considered (low capability peers cannot collaborate as much as high capability ones).

Furthermore, a contributing peer timelessly participates at increasing system's usefulness and resource availability and/or diversity.

They are equally increased by staying on and by letting shared resources in the resource pool.

Resource availability. When multiple peers hold an item of a replicable resource, its availability is increased. The workload is thus better balanced among and response delays decreased.

Resource diversity. Whenever peers contribute with new resources to the system, resource diversity is increased. However, offering such a resource may make demands converge on its provider and may cause congestion.

Hence, resource abundance keeps the system from the tragedy of commons as peers do not compete against each other to access a resource. Since many peers provide a resource, requesting peers can target different providers and get satisfied more quickly.

B. Incentive mechanisms roles

An incentive mechanism must assume the following roles.

Conditioning access to the resources (dealing with newcomers),

Detecting and sanctioning non-contribution,

Compensating contribution cost,

Preventing peers from escaping sanctions or the cost of sharing,

adapts to peer behavior change and peer's ability for providing resources.

An incentive mechanism must deal with newcomers and sets the entry rules to the system. A system with a free entry might dissuade peers from participation and make non-contribution a possible option.

On one hand, a non-contributor is a peer that does not propose any resource while on the other hand, a defector, advertises resources but does not effectively share them when asked for.

The system must distinguish contributor and non-contributor peers. To detect non-contribution, collaborative operations watched by the system must be enumerated. A peer's participation level is defined in order to classify the peer as collaborative or non-collaborative for each of the considered operations. A global definition of peer's

participation level considers then each of the watched collaborative actions and the QoS required for acceptable performances of the system.

Contribution should be the only alternative for the peer to access resources. Costs compensation can be immediate or delayed. Incentive factors must be set and disincentive factors reduced or eliminated.

An incentive factor is compensation for the cost of sharing whereas disincentive factors are all what can be exploited by peers not to contribute but still benefit from the system.

A peer is punished by being temporarily disabled to access resources. Peer's behavior might change. The incentive mechanism should consider redemption: the peer that contributes again gets access to resources again. Punishment depends on the targeted system and application types. It can be restricting resource access or exploiting the peer by increasing the peer's workload.

All peers cannot contribute the same resources, then, peer's capability must be taken into consideration.

III. INCENTIVE MECHANISMS IN P2P SYSTEMS

Incentive mechanisms vary according to the specification level (generic, specific to a system's type or an application) and to the incentive scheme. The specification level allows a better view of collaborative operations in the system and a definition of non-contribution.

A. Generic incentives

Generic incentive mechanisms are abstract from the system's type and only consider resources and peers. Most of them are theoretical incentive models featuring economic theory (game theory and mechanism design) where P2P systems are viewed as a set of actors (peers) that base on their own information, capabilities and preferences to behave in a self-interested way while ignoring system welfare. Combined peers behaviors lead to a certain system's state. These propositions study the behaviors of peers with or without incentive mechanisms and are analyzed to show the evolution of P2P systems with and without incentives. Effectiveness and global optimality at equilibrium of the theoretical models must be proven.

These propositions may provide general guidelines for practical incentive mechanisms design. We argue that if a generic incentive mechanism can capture most important P2P systems' characteristics, they do not consider the proper functioning and the performance metrics of a particular application. They might also not be robust enough against specific misbehaving patterns of an application. Specification of incentive mechanisms guarantees more efficiency. Various game models of game theory are used from which we cite [3] for evolutionary prisoners' dilemma (EPD), [4] for generalized prisoners' dilemma (GPD) and finally repeated games in [5].

B. Specific incentives : targeting specific systems and applications

Impact of non-contribution and effectiveness of incentives depend on P2P system and application types.

Some incentive mechanism propositions specifically encounter misbehavior in centralized, decentralized structured and unstructured P2P systems and might be

application specific. Indeed, contributors and non-contributors are identified with regard to the considered collaborative operations.

Observed collaborative operations consume peer's physical resources (bandwidth and storage space) and are as follows.

Resource lookup. Without incentives, search can be degraded.

- In unstructured P2P systems, peers might not transfer packets to neighbors.
- In structured P2P systems, peers might neither hold nor maintain routing tables nor route request packets.
- In a hybrid P2P system, the super-peer informs affiliate peers about resources location in the system. A peer might not assume the super-peer role.

Resource hosting in structured P2P systems. In structured file sharing systems, files are relocated at peers with the closest identifier. This supposes that the host peer is willing to give up some memory space for storing the file and bandwidth when responding to requests.

Effective resource sharing. In advertisement based P2P systems, resource advertising does not guarantee effective sharing and cannot be considered contribution.

Information providing. Some protocols require peers to provide others truthful information for system's functioning.

Observed collaborative operations can also be *applicative*. Peers are considered untruthful and able to defect in every possible application specific collaborative task as they can modify client software code.

The target applications are content distribution applications (file sharing and multimedia streaming applications), distributed storage or distributed computing applications. According to the applicative goal, the mainly solicited physical resources are different.

1) *Content distribution systems*

A copy of the requested content is sent to the requesting peers. Transferring content temporarily involves sharing provider's bandwidth. In multimedia streaming systems, stream transfer delay must be the shortest possible as streams are viewed in real time; they thus require greater amounts of bandwidth than in file sharing systems.

2) *Distributed storage*

Distributed storage involves similar resources than file sharing networks except that peers must be guaranteed to be able to retrieve their data when needed.

3) *Distributed computing*

Peer's processing power is the primary resource. Memory and bandwidth are queried for storing intermediary results and getting in and out input and output data respectively.

Measuring physical resources contribution is done by considering the amount of the most critical resource of the application. However, measuring the resource contribution is only necessary if the peer is satisfied of the service that it has

been offered (for example, in file sharing systems, the peer should be satisfied of the received file in terms of its completeness, its integrity and its quality).

C. *Incentive schemes*

There are two types of incentive schemes: economic based and reciprocity based.

1) *Economic based incentives*

Economic incentives create an economy where providing a service is lucrative. A peer has to pay the contribution of other peers with regard to the considered collaborative operations. The remuneration (virtual coins) is exchanged between transacting peers. The peer does not consume a resource unless it has contributed before and owns sufficient amount of currency to pay for the resource. The only behavior that shows is contribution: non-contribution does not show to other peers but still affects the peer that does not have the required amount of currency to get resources.

The drawbacks of such mechanisms are the important overhead imposed by the communications and the requirement for persistent and strong peer's identification [6]. Persistent identification means that the identifiers are valid over a long period of time. Strong identifiers are resistant to peers' malicious intentions like impersonating. They also may have a single point of failure if a function is implemented in a centralized way. Indeed, they might need a central authority to distribute the currency. Finally, security is needed in payment schemes to ensure robustness of the incentive mechanism.

Economic based incentives differ from each other depending on: the observed operations, the resource pricing scheme, who is paid, and the payment mechanism.

The resource provider is paid for providing the resource. Intermediary peers participating in the transaction may get paid or not depending on whether routing packets is an observed operation.

The resource price can be fixed or variable over time but must be fair in the sense that it must reflect the real cost of resource provision. A trusted entity sets the price in fixed-price schemes whereas in variable-price schemes, the price can either be fixed by a trusted entity or by one of the transacting peers. On one hand, if the price is settled by the peers providing resources, it should reflect the availability of resources the peers share and their workload (which depends on owned resources' popularity). Provider peers might also be tempted by overpricing the resource. In case where the resource's availability value in the system is high, the requester get to choose among the several prices' propositions which places the peers offering the requested resource in a situation of competition. On the other hand, if requester peers settle the price, they will tend to lower it. Another option is that several requesting peers targeting a resource at the same provider bid for it.

Payment mechanisms [7, 8] specify coin valuing (the value of coin with regard to one unit of provided resource), coin issuing responsible at the bootstrapping phase, the payment strategy (when the resource is paid: before transfer, during transfer, after transfer) and the actual coin transfer method (accounting). It should detect double spending, coin fraud and any possible malicious attack. Moreover, payment transactions have to be controlled and peers identification is needed for accounting so that each peer has its account and

cannot cheat the system. Both centralized and distributed implementations are possible for a payment mechanism.

Authors in [10] propose an incentive mechanism where both provider and routing nodes are paid for providing and transferring the file. A file is supposed to be held by only one server. There are n disjoint paths from the client to the server and peers at one hop from the server are called *terminal nodes*. For a wanted file, a peer has a utility and a maximum price offered P_C . It sends lookup messages through the different paths and each intermediate node adds its marginal cost to the total marginal cost. An intermediate node's marginal cost depends on its cost of forwarding the request (it is higher if the peer is loaded). Terminal nodes submit a bid (for resource payment) that equals the offered price minus the total marginal cost. The server runs a Vickrey auction where the highest bid path wins the auction and pays the second offered price. The profit is then fairly shared between the nodes of the winning path in proportion to their marginal cost. The auction is secured to avoid cheating. We argue that authors were not realistic when supposing that a resource is offered by a unique server and the paths to it are disjoint. In a P2P file sharing network, a file might be offered by multiple peers and finding disjoint paths from a client to server is costly.

[9] addresses the dropping problem. Authors suppose resource abundance and a pay-per-service system. Some peers provide the same resources with different prices and are in competition with each other. Thus, if they do not forward queries, a peer offering a less pricey resource does not get chosen. This incentive mechanism tries to insure a proper search mechanism. The protocol is based on *RTR(Right To Respond)* messages. *RTR* messages are encrypted. A peer must pay to receive a request as responding to it may bring revenue. Once the *RTR* is bought, the peer has two options: to respond to the request if it owns the corresponding file and hopes to be chosen to provide the service or to sell the *RTR* to another peer. Before a peer A forwards the query to a peer B , it sends it an offer message; this latter is necessary to the peer B to decide whether to buy the request or not. This proposition lacks details about pricing scheme and payment mechanism.

2) Reciprocity based incentives

One peer might decide to collaborate depending on its knowledge of others present and/or past behaviors. A peer offers its resources to contributors and defects on non-contributors.

A common issue for reciprocity schemes is dealing with newcomers. Newcomers might be hard to transact with especially if they have no initial resource to contribute. They might be dissuaded from joining the system if joining rules are too restrictive or enticed to whitewashing (escaping consequences of past non-contribution) if the joining rules are too loose. A fixed-fee entry rule or a stranger adaptive strategy might be established. The fixed-fee entry rule conditions access to the system by a specified amount of resources to provide. The stranger adaptive strategy states to treat strangers according to the behavior of the past newcomers. Another solution to whitewashing is assigning persistent identifiers for all newcomers.

Another problem is peer's identification. A peer's identity must be strong and persistent over time (the longest possible). Whitewashing is feasible if identities are not strong. Peers should not be able to escape their past

misbehavior by entering the system under a new identity and allow a coherent information keeping.

Reciprocity based schemes are divided into two subclasses: direct reciprocity schemes and indirect reciprocity schemes.

a) Direct reciprocity schemes

Direct reciprocity schemes consider interactions between each couple of peers.

On one hand, a peer A provides resources to a peer B if B has already provided resources to A , B is providing resources to A , B is going to provide a resource to A . On the other hand, a peer A might consider the peer's B priority among other asking peers according to their contribution.

Drawbacks of these mechanisms are that the only possible operation to consider is resource provision and the need of mutual interest for owned resources.

These mechanisms can be further classified according to the reciprocity time into *immediate* and *deferred* direct reciprocity schemes.

Immediate direct reciprocity schemes

In immediate ones, resource exchange between the interacting peers is simultaneous. After an agreement on the exchanged resources, the transaction phase starts. Transacting peers do owe nothing to each other at the end of the transaction.

The advantages of immediate direct reciprocity schemes are as follows:

- no need for keeping history of the previous interactions with peers,
- no need for strong and persistent peers' identification mechanism,
- little overhead,
- it can detect freeriding in real time and react by stopping the current resource provision,
- no need for a central entity to track the reciprocity.

The drawbacks of this type of mechanisms are that peers need an immediate mutual interest for the currently owned resources (so that they can immediately exchange them), non-scalability and a possible delay before the start of resource provision.

While this scheme could be interesting to implement, to our knowledge, there have been no proposition.

Deferred direct reciprocity schemes

Deferred direct reciprocity schemes differ from the immediate ones in that compensation for providing resource is not immediate: there is no initial agreement on the exchanged resources before the first resource transfer.

Their advantage is that immediate mutual interest for owned resources is mandatory.

The drawbacks of deferred direct reciprocity schemes are:

- the possibility of a peer leaving the system after getting the resource,

transaction information must be kept for future reciprocation,

long duration sessions are required for repetitive interactions: this is not always possible in reality and choosing a peer to interact with becomes harder.

Authors in [11] propose *SLIC*, a direct deferred reciprocity based incentive mechanism for unstructured P2P networks. Each peer has a capacity (that it uses either for generating new queries or responding to received queries), an answering power (the probability of having the asked for resource) and a set of neighbors. Every peer monitors its neighbors in the operation of processing and forwarding queries, thus, peers are guaranteed better service if they provide more capacity to respond to requests and if they establish more connections. A peer periodically computes its neighbors weights (its opinions about its neighbor), whose values vary from 0 to 1: 0 means that the neighbor is useless and 1 means that it is an excellent neighbor. A peer gets better service if its neighbors give it a higher weight value. Weights computation takes into account the query hit messages neighbors originate during a period for the queries with expired *TTL*. The peer then decides if its neighbor behaves well enough for processing or forwarding its requests. The paper also addresses the initialization of a newcomer (to avoid whitewashing and strict restrictions to enter the network) with three possible schemes. Finally are discussed two possible options to decide if peer should accept a connection request of a newcomer unconditionally: limited or unlimited degree.

b) *Indirect reciprocity schemes*

Indirect reciprocity schemes consider interactions between more than two peers.

Advantages of indirect reciprocity schemes are that frequent transactions between two peers are not necessary and the scalability to large populations in comparison with direct reciprocity schemes. Their drawback is that they may rely on second hand observations which raise trust issues between peers.

Indirect reciprocity schemes are *exchange based* or *reputation based*.

Exchange based reciprocity schemes

In an exchange based reciprocity scheme, the peers organize in groups (a ring or a swarm) in which they share resources in an immediate way. In a ring, there is an organized transitive resource exchange: a peer is served by its predecessor and serves its successor. A swarm is a set of peers where a peer might serve and might get served by multiple peers at the same time. It gathers peers with the same interests. If the resource is dividable (a file composed of chunks), the swarm is the set of peers sharing chunks of that same file.

The advantages of such schemes are that resource sharing is stimulated and delay latency for resource getting is reduced.

Their drawbacks are that mutual interest in resources is needed to form groups and group formation is a dynamic process that induces more resources consumption.

In [12] is proposed a bartering based approach where peer's received service quality is proportional to the one it provides. A peer downloads while uploading. The uploading

bandwidth in the ring is the smallest bandwidth offered by a peer. Peers are encouraged to increase their uploading bandwidths in order to see their download speed increasing. In our view, however, authors do not give many details about rings construction and its feasibility. This method totally eliminates freeriding if cheating (a peer pretending that it has file but does not) is not considered possible. Rings construction increases the delay before starting downloads and imposes a communication overhead because of exchanged messages.

In BitTorrent [13], the freeriding problem is addressed with the tit-for-tat (TFT) incentive strategy that entices peers to seed. However, through implementing a strategic BitTorrent client, authors in [14], [15] have proven the inefficiency of this strategy to provide robustness: BitTorrent's well functioning is not due to TFT but to some altruism in the system. Freeriders are still able to cheat the system and achieve high quality downloads (optimal downloading speed and time). Still, with the TFT strategy, a seeder gains nothing in comparison to leechers or freeriders.

Reputation based reciprocity schemes

In indirect reputation based reciprocity schemes, a peer relies on reputation value of another peer to decide whether to transact with it or not. Reputation value is computed using its local information about transaction outcomes and/or some or all others opinions. The reputation value represents the trust that a peer has towards another and reflects both peer's capacity and motivation. Negative feedback translates into a lower reputation value while positive feedback into higher reputation value. On one hand, the providing peer might decide whether or not to provide the resource based on the reputation value of the demanding peer. On the other hand, the requesting peer might get refused resources because of a poor reputation. The peer gets enticed to behave well in order to build a good reputation and get served.

A reputation mechanism assesses a reputation value (trust degree) based on a contribution level and has five functions:

- information gathering,
- reputation value assessment,
- reputation value storage,
- reputation value dissemination,
- the decision function.

Information gathering. After each interaction, the peer getting the service reports the transaction's success or failure and its satisfaction value of the observed collaborative operations based on the expected QoS and the received QoS.

Information sources can be restrained or not.

Reputation value assessing. The aggregation of satisfaction values into a reputation value is specific to each reputation mechanism and its computation obeys to a method. When to compute the score is another factor of reputation mechanisms as the behavior varies through time. A frequent update might not be necessary as the computation gives the same value whereas a less regular one could lead to cheating: a peer with a high enough reputation can stop contributing and still benefit from it. A solution is considering short term history [16]. Short term history

ensures that a peer that was once a good contributor and that has not been anymore in a nearest period of time does not benefit from its previous good reputation.

The score value type depends on its evaluation: it can be binary (1 for contributor, 0 for non-contributor) or real and comprised between 0 and 1.

Reputation is also classified into *autonomous* (relative) or *global* (absolute) reputation. In autonomous reputation schemes, the peer relies on what it knows from its transactions with other peers; values are neither disseminated nor aggregated. In global ones, information is disseminated to and aggregated from multiple or all peers.

Reputation value storage. A centralized storage induces a single point of failure. The distributed storage makes the reputation mechanism more vulnerable to peers misbehavior. The responsible of computation can either send the reputation value on demand to a specific peer, on update or periodically or send a blacklist of guilty peers to all peers.

Reputation value dissemination. Reputation mechanisms differ from each other depending on how reputation values are disseminated, to whom they are disseminated and when they are disseminated. If the reputation value is sent through specific messages, the respective mechanism is *explicit*. If reputation value is adjoined to a message, the mechanism is a *gossip* mechanism [6]. It could be sent on update or on demand. It is sent on demand to the peer that asked for it or on update to all systems peers or a restrained group of peers.

The decision function. The incentive mechanism maps the reputation value into a type of peer (collaborative or non-collaborative) and then to reward or sanction.

As reward can be proportional to the contribution level and the quality of the offered services, punishment can be proportional to the harm the peer caused to the system.

The advantage of reputation based mechanisms is that direct reciprocity in resource providing between peers is not necessary.

Their drawbacks are as follows.

Peers unreliability in information reporting.

Information quality and correctness of reputation values.

Reputation mechanisms are subject to individual or collective false reporting (false accusation or false praise). Information validity and integrity are not ensured and consciously sharing incorrect information about peer's reputation should be sanctioned. Sybil attack is a possible individual malicious behavior where a user joins the system under different identities and uses them to enhance the reputation of one of them. Collusion is an example of collective malicious behavior where a group of peers declare false transactions in order to improve each others' reputation.

Correctness of reputation values depends on the gathered information quality in terms of *completeness*, *accuracy* (*timeliness*) and *truth*.

Information completeness depends on the number of information sources. Ideally, a peer *A* gets feedback about a peer *B* from all other peers that have already interacted with it. Restraining information sources to a set of peers (in the local neighborhood) implies integrity issues and potential group attacks (collusion). Increasing information sources decreases the quality of each of them [17]. Not restraining

information sources allows an accurate evaluation but can lead to additional overhead. Moreover, the gathering time influences information correctness.

Information accuracy depends on whether the information is up-to-date or not. Accuracy might also consider freshness. Recent peer's behavior might be completely different from past peer's behavior. Defining a window allows to consider the most representative past behavior.

Truthful information revealing should be enticed. Reputation values of sources can be used for weighing their opinions in handling information integrity.

Good information quality implies high overhead: a compromise between the three previous factors is thus necessary.

Two kinds of reputation mechanisms exist: *centralized* and *decentralized* [18]. In centralized implementations, an objective central authority is responsible for gathering, computing, storing and disseminating reputation values. In decentralized ones, all peers or some trusted peers gather information, compute, store and propagate reputation values. Another possibility is that only peers that have been involved in transactions with other peers are actors of the reputation mechanism. This latter option imposes the necessity of coherence of reputation values maintained at each entity.

While the impact of bad behavior should be superior to that of good behavior [17], indirect reciprocity schemes do not consider non-contribution level to compute reputation values, and cannot be perfectly representative since defection motivations are not known.

Authors in [19] propose a reputation based incentive mechanism for P2P file sharing networks where a peer is enticed to share files while downloading. At joining, a peer must advertise a minimum number of files and effectively share them when asked for. A decentralized index holds the peers' shared files and contains a field where a peer can check the validity of the resources provided by the peer it is serving. A peer is not served if it refuses to serve files to others. When the peer *A* wants to download a resource from the peer *B*, *B* is interested in knowing if *A* is sharing the files it advertises. If *A* refuses to share a file to a peer *C*, this latter informs *B* (it gets its identity from the index) and *B* stops uploading to *A*. The same arrives if the shared file is not valid. In order to increase availability time, the average download time is increased by fixing the upload bandwidth: the peers' presence time in the system is longer.

Authors in [20] propose a middleware of four components to stimulate fair cooperation. A peer joins the system by sharing a set of resources it is willing to contribute. It is assigned a strong identity by the *registration service* to avoid whitewashing and an initial access level reflecting the portion of resources it has access to as a newcomer. The access level value evolves proportionally to participation level and both are computed by the *aggregation service*. A peer's participation value is increased when it responds to a query and decreased if it forwards a query while it should have responded to it. Fairness in access to reward is ensured in the following way: if two peers are able to respond to a query, it is the one with the smallest participation level that responds to it (and thus gets the chance to improve it). *Semantic group membership*

service allows grouping peers with common interests (they share at least one resource of the type of other peers in the same group: semantic groups augment the potential of sharing between peers) and supervising them by supervisors (trusted peers): peer's queries are responded by neighbors in semantic groups for each type of resource. Peers are tracked when their suspicion level reaches a certain threshold by the *tracking service* that prioritizes them for responding to request. If they do not, after a time-out, they are notified freeriders to the tracking service with a probability of the requester's access level. Peers also notify the outcome of their transactions to this service. When participation level increases, suspicion level decreases. Authors, however, do not consider trust issues and information misreporting.

Authors in [21] deal with reporting truthful feedback information in P2P systems. Peers involved in a transaction have to rate each other on the transaction quality and submit a feedback message to *credibility holders*. Credibility holders are peers structured in a DHT that are responsible for storing and providing reputation values for transacting peers. If peers disagree on a transaction quality, one of them is lying and both should be punished by a temporary ban from transaction. Duration of punishment depends on and is proportional to the *non-credibility* value. Each peer should send feedback messages to every credibility holder maintaining credibility information about the transaction. Newcomers are assigned a default value of non-credibility. Each peer has a pair of public and private keys to sign feedback messages. Receiving a feedback message from only one transacting peer is considered as disagreement and is punished. In order to ensure truthfulness of credibility holders, a peer periodically asks for its non-credibility values from credibility holders and compares them. This trust mechanism allows a better peer selection as peers do compute reputation values are influenced by non-credibility ones. One drawback is that it still cannot distinguish truthful peer from the untruthful peer and hence punishes the innocent peer.

IV. DISCUSSION

In a P2P system, the incentive mechanism is responsible for identifying and stimulating non-contributors to provide resources while compensating contribution. Non-contribution definition depends on the system functioning and the quality of service required from the system.

Targeting a specific application type or system adds to the mechanism's efficiency as it allows an accurate definition of contribution (observed collaborative operations) and non-contribution (setting the thresholds under which the peer is considered non-contributor).

Every proposition of a practical incentive mechanism requires a specification of the target system and/or application. Opting for an incentive scheme requires a compromise between:

- researched system's performance and how observed collaborative operations affect it,
- advantages and drawbacks of incentive schemes and their implementation costs within the specific system.

Incentive mechanism's functions are undertaken in a centralized, decentralized or hybrid way. We argue that opting for one of them depends on the specific needs of the

system. Compromises are to be made between advantages and drawbacks of centralization and decentralization.

We can compare the incentive mechanisms on the following parameters: the targeted system and application, the observed operations, the incentive strategies and the type of incentive mechanisms.

Incentive mechanisms requirements are scalability, robustness to peers threatening or escaping consequences of their behaviors and a minimum communication and information storage.

Incentive mechanisms design and implementation confront the challenge of security and peer's identification. The frontier with security issues is very subtle and incentive mechanisms must define what is considered as a valid contribution: it should be provision of an interesting and useful resource. Incentive mechanisms efficacy is closely related to peer's identification mechanism. If a peer can escape the incentive mechanism constraints or penalties, the mechanism is no longer effective. They also have to resist malicious attacks and be resilient to failures. Attacks range from individual (false praise, treachery, whitewashing and Sybil attacks) to collective attacks (like collusion).

False praise: sharing false values.

Treachery: a peer may behave well and change strategy while still benefiting from its past good reputation.

Sybil attacks: a same peer appears under multiple identities in order to influence a collective decision by submitting a choice several times.

Collusion: peers group themselves so they can make transactions with each other in order to enhance their reputation value.

V. CONCLUSION AND FUTURE WORK

This paper addressed incentive mechanisms in P2P systems. We first highlighted roles and objectives of an incentive mechanism. We then classified incentive mechanisms according to their specification level and the incentive scheme they implement. Generic incentive propositions are exploited to design more specific incentive mechanisms targeting a specific system and/or application type. Specifying the system and application type allows a precise definition of contribution and hence a clear identification of non-contributors according to the chosen observed collaborative operations. Once observed operations and contribution assumptions clearly set, incentive schemes are to be chosen according to the system's overall needs in terms of functioning and performance. Major challenging issues related to incentive mechanisms are security and peers identification. Other issues target more specifically some incentive schemes rather than other like defining efficient pricing or payment mechanisms in an economic incentive scheme. We also underline the features, advantages and drawbacks of each incentive scheme. Finally, we argue that this classification is helpful in incentive mechanism proposition targeting static or mobile P2P systems.

REFERENCES

- [1] M. R. Rahman, "A Survey of incentive mechanisms in Peer-to-Peer systems," 2008.
- [2] S. B. Guruprasad Khataniar and A. Khataniar, "A comparative study of incentive mechanisms used in Peer-to-Peer system," 2012.

- [3] K. Lai, M. Feldman, I. Stoica, and J. Chuang, "Incentives for cooperation in peer-to-peer networks," 2003.
- [4] M. Feldman, K. Lai, I. Stoica, and J. Chuang, "Robust Incentive Techniques for Peer-to-Peer Networks," in Proceedings of the 5th ACM Conference on Electronic Commerce, ser. EC '04. New York, NY, USA: ACM, 2004, pp. 102–111.
- [5] H. Chen, H. Xu, and L. Chen, "Incentive mechanisms for P2P network nodes based on repeated game," JNW, vol. 7, no. 2, 2012, pp. 385–392.
- [6] M. Karakaya, I. Korpeoglu, and O. Ulusoy, "Free riding in Peer-to-Peer networks," IEEE Internet Computing, vol. 13, no. 2, 2009, pp. 92–98.
- [7] B. Yang and H. Garcia-Molina, "PPay: Micropayments for Peer-to-Peer systems," in Proceedings of the 10th ACM conference on Computer and communication security. ACM Press, 2003, pp. 300–310.
- [8] F. D. Garcia and J.-H. Hoepman, "Off-Line Karma: A decentralized currency for Peer-to-Peer and grid applications," in ACNS, ser. Lecture Notes in Computer Science, vol. 3531, 2005, pp. 364–377.
- [9] B. Yang, T. Condie, S. Kamvar, and H. Garcia-Molina, "Non-cooperation in competitive P2P networks."
- [10] R. Gupta and A. K. Somani, "Pricing strategy for incentivizing selfish nodes to share resources in Peer-to-Peer (P2P) networks," vol. 2, 2004, pp. 624–629.
- [11] Q. Sun and H. Garcia-Molina, "SLIC: A selfish sink-based incentive mechanism for unstructured Peer-to-Peer networks," in 24th International Conference on Distributed Computing Systems, 2004, pp. 506–515.
- [12] T. Ackemann, R. Gold, C. Mascolo, and W. Emmerich, "Incentives in Peer-to-Peer and grid networking," UCL-CS, vol. 1, 2002, pp. 1–12.
- [13] B. Cohen, "Incentives build robustness in BitTorrent," 2003.
- [14] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do incentives build robustness in BitTorrent," in Proceedings of the 4th USENIX Conference on Networked Systems Design & #38; Implementation, ser. NSDI'07. Berkeley, CA, USA: USENIX Association, 2007, pp. 1–1.
- [15] S. Jun and M. Ahamad, "Incentives in BitTorrent induce free riding," in Proceedings of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems, ser. P2PECON '05. New York, NY, USA: ACM, 2005, pp. 116–121.
- [16] M. Feldman, K. Lai, I. Stoica, and J. Chuang, "Robust incentive techniques for Peer-to-Peer networks," in EC '04: Proceedings of the 5th ACM conference on Electronic commerce. New York, NY, USA: ACM, 2004, pp. 102–111.
- [17] S. Marti and H. Garcia-Molina, "Taxonomy of trust: Categorizing P2P reputation systems," Computer Networks, vol. 50, no. 4, 2006, pp. 472–484.
- [18] A. Blanc, Y.-K. Liu, and A. Vahdat, "Designing incentives for Peer-to-Peer routing," in INFOCOM. IEEE, 2005, pp. 374–385.
- [19] P. Antoniadis, C. Courcoubetis, and B. Strulo, "Incentives for content availability in memory-less Peer-to-Peer file sharing systems," SIGecomExchanges, vol. 5, no. 4, 2005, pp. 11–20.
- [20] E. Anceaume, M. Gradinariu, and A. Ravoaja, "Incentives for P2P fair resource sharing," in Peer-to-Peer Computing. IEEE Computer Society, 2005, pp. 253–260.
- [21] T. G. Papaioannou and G. D. Stamoulis, "An incentives' mechanism promoting truthful feedback in Peer-to-Peer systems," in CCGRID, 2005, pp. 275–283

Optimisation du routage des navires par une approche hybride basée sur l'ACF

Ibrahim HENNI¹, Khaled BELKADI²

LAMOSI, USTO-MB Université Mohamed BOUDIAF
ORAN, ALGERIE

¹ hennibrahim@live.fr

² belkadi1999@yahoo.com

Résumé—Face à la hausse du prix international du pétrole et le réchauffement climatique, les concepts du «transport vert» ont été soulevés dans l'industrie du transport maritime international. Par conséquent, la gestion du transport doit prendre soin non seulement de l'efficacité économique, mais aussi de la protection de l'environnement pour prévenir la pollution. Les Systèmes d'Information Géographique (SIG) ont une place prépondérante dans le cadre de la navigation maritime. Les déplacements des navires dans un environnement maritime sont planifiés. La planification efficace de route empruntée par un navire est généralement déterminée par trois aspects : la précision de la prédiction du comportement hydrodynamique du navire dans différentes conditions météorologiques, l'exactitude des prévisions météorologiques, et l'algorithme d'optimisation. Dans ce papier nous avons apporté des modifications à l'Algorithme de Colonie de Fourmis (ACF) pour l'application de la navigation maritime, et nous avons intégré les concepts de l'Algorithme Génétique (AG) pour améliorer son efficacité.

Mots clés— *Métaheuristiques; Optimisation multiobjectif; SIG; Route Maritime Optimale; Système de navigation*

I. INTRODUCTION

Dans le monde globalisé d'aujourd'hui, le transport de marchandises en mer est d'une grande pertinence: plus de quatre-vingt dix pour cent (90%) du commerce mondial est transporté sur mer [1]. Selon les statistiques, le volume de marchandises transportées par voie maritime a eu des effets négatifs importants sur l'environnement, ce qui peut causer un réchauffement climatique et une pollution de l'air. Par conséquent, une raison supplémentaire pour réduire la pollution de l'air est de diminuer la consommation de carburant pendant la navigation [2].

Le choix de la route optimale, le maintien de la vitesse et l'évitement des obstacles peuvent non seulement réduire la consommation de carburant de façon significative, mais aussi améliorer la sécurité de la navigation. Un équilibre entre la sécurité de la navigation et de l'efficacité économique doit être trouvé [3] parce que la détermination de la route optimale est un problème multicritère non-linéaire et englobe de nombreuses contraintes. En d'autres termes, le processus de la planification ne doit pas prendre en compte seulement l'évaluation des risques sécuritaires en cas de déviation pendant la navigation, mais aussi éviter une augmentation du coût total en raison des mesures de déviation.

Le problème de planification de routes maritimes consiste à minimiser le temps de parcours et les distances parcourues d'un navire pour aller d'un point courant à un objectif. Diverses contraintes météorologiques peuvent être considérées : force du vent et caractéristiques des vagues. On

s'intéresse au problème de navigation maritime. Il est connu dans la littérature comme étant NP- Difficile [4]. Pour résoudre des instances de taille importante, une solution consiste dans un premier temps à faire recours aux méthodes d'optimisation combinatoire telles que les métaheuristiques.

Dans ce papier, nous allons étudier le problème du routage optimal des navires en modélisant le déplacement et en dégagant une solution basée sur un algorithme d'optimisation qui optimisera les déplacements effectués par les navires et qui tiendra compte à la fois de la sécurité et de l'économie. Afin de réaliser cette mission d'optimisation, nous avons proposé un nouvel Algorithme de Colonie de Fourmis (ACF) avec l'intégration des concepts de l'Algorithme Génétique (AG). Le calcul et la recherche de l'information géographique sont fournis par le Système d'Information Géographique (SIG) développé.

II. NAVIGATION MARITIME

La navigation désigne l'ensemble des techniques et méthodes qui permettent au navire de déterminer sa position et de calculer la route à suivre pour se rendre à sa destination, en toute sécurité [5].

A. Route maritime

Il n'existe pas aujourd'hui de données répertoriant les routes réellement suivies par les navires parcourant les mers du globe. On dispose par contre des ports de départ et d'arrivée de la plupart des voyages commerciaux, d'informations ponctuelles sur la position de certains navires à proximité des côtes, ainsi que de données sur les grandes routes généralement empruntées par les navires.

Aujourd'hui, dans l'évaluation des routes maritimes optimales, la principale préoccupation, outre la sécurité, est généralement accordée aux temps de navigation (ou vitesse de navigation) plutôt que de la distance. En considération des critères de la sécurité et de la consommation d'énergie, la plus courte distance entre deux ports n'est pas nécessairement une route optimale. Il ya de nombreuses routes à choisir entre elles, la route qui prend le moins de temps est la route optimale.

B. Calcul de la route optimale

La résolution d'un problème de route optimale consiste à trouver la route avec le moindre coût parmi l'ensemble des chemins candidats, en tenant en compte les conditions météorologiques. Le coût mentionné n'est pas limité à la distance seulement, mais doit prendre en considération le temps de navigation, la consommation de carburant, ou une combinaison de facteurs. Les variables de décision qui doivent être contrôlées sont les manœuvres de cap et de vitesse. Le coût total (I) est calculé comme la somme des

coûts sur chaque segment de route (s). Il est basé sur la position du navire (P), des changements dans le contrôle de manœuvre (C) (i.e., vitesse et cap) et la durée en temps (t). Avec ces facteurs de coût, le problème de routage des navires peut être décrit par la formule suivante:

$$I = \int_S f(P(s), C(s), t(s)) ds \quad (1)$$

où :

- f (P, C, t) fonction du coût du voyage pour une position donnée, contrôle de manœuvre et un temps t; et $P \in \mathbb{R}$, $C \in C_A$.
- R une région possible de navigation.
- C_A variable de contrôle de direction du cap et de maintien de la vitesse.
- ds est mesuré par la consommation de carburant et le degré de sécurité des navires navigants.

C. Modélisation des trajets

Les progrès dans la prise en compte de la connaissance en temps-réel des informations sur les déplacements des objets mobiles s'accompagnent du développement des systèmes d'informations géographiques (SIG) qui favorisent, pour leur part, la gestion et la manipulation de ces données.

L'équipe d'Yvan Bédard de l'université Laval au Québec a développé une méthode générale qui permet d'étendre n'importe quel formalisme graphique de modélisation pour la conception géographique. Cette méthode est une extension des diagrammes de classes UML pour la définition des données spatio-temporelles. Plusieurs notations ont été reprises de la méthode entités-relations Modul-R[6].

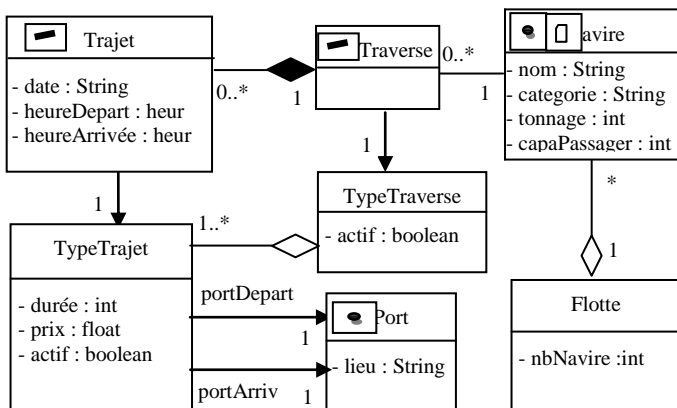


Figure 1 : Schéma conceptuel GEO-UML

III. ALGORITHME DE COLONIE DES FOURMIS ET LE ROUTAGE DES NAVIRES

Le principe s'inspire du comportement de masse de colonies de fourmis. Dans le monde naturel, dans la recherche de la nourriture par des fourmis, elles prennent les différents chemins possibles du nid à la source de nourriture, différentes fourmis vont d'abord choisir des chemins différents mais à la fin, la plupart des fourmis vont découvrir le même chemin le plus court. Cette recherche du plus court chemin est un processus interactif.

A. Le principe de l'algorithme des colonies des fourmis (ACF)

On met un certain nombre de fourmis artificielles (navires) qui suivent un ensemble de règles de recherche pour trouver la route optimale globale. L'objectif de routage

des navires est de trouver la plus sûre et la plus économique route de navigation basé sur les conditions météorologiques et selon les règles d'évitement des plates-formes pétrolières fixées au fond marin entre les points de départ et de destination.

B. Division du système de grille marine

Afin de planifier les changements de cap nécessaires dans la route de navigation, l'espace maritime doit d'abord être divisé en grilles [7], et la route de navigation se compose de *Waypoints*. La procédure pour faire une grille est la suivante:

- Établir d'abord une route orthodromique entre le point de départ ($L1, \lambda1$) et le point final ($L2, \lambda2$), pour servir de base à la planification de la route.
- Ensuite définir une distance X, et définir les points des segments le long de la route orthodromique ($Lx, \lambda x$) dans des intervalles de X.
- Puis connecter les points de segmentation dans le Nord-Sud, Est-Ouest pour former un système de grille.
- Enfin définir la matrice des coordonnées et de nœuds, et la planification de la route sera établie dans ce système de grille.

Avec les avancées de la communication et de la technologie de positionnement, le système de grille et la route de navigation peuvent être ajustés de manière dynamique et continue, selon les dernières conditions météorologiques, le dernier point de positionnement et le point final sur la route orthodromique.

C. L'implémentation de l'algorithme des colonies des fourmis

La première étape consiste à établir un système de grille et sa matrice de nœuds dans la zone de navigation, et fournir à chaque nœud avec la valeur initiale appropriée dans la matrice pour former la matrice initiale des informations de nœud. Ensuite, mettre toutes les fourmis sur le point de départ pour leur permettre de se déplacer simultanément vers la direction de l'objectif, et pour atteindre le point de destination à la fin. Dans le processus, chaque fourmi utilise la règle de transition d'état et soumis à un processus de sélection lors du passage au prochain nœud. Toutes les fourmis vont atteindre la destination simultanément et compléter une route. Après que toutes les fourmis ont atteint la destination, calculer la valeur de la fonction objectif en se basant sur des routes découvertes par chaque fourmi. Chaque nœud est également mis à jour avec la quantité globale de phéromone, chaque fourmi k laisse une quantité de phéromone dans chaque arc qu'elle choisit. La mise à jour renforce les arcs choisis en prenant en considération l'évaporation. Ce processus sera répété jusqu'à ce que la route optimale soit découverte. L'explication détaillée du processus est la suivante:

1) Réglage de la quantité de phéromone

Le nombre de fourmis est supposé égal à m . Et, chaque fourmi choisit la route suivante en fonction de la probabilité correspondante sur la base de la quantité de phéromone sur le trajet, au lieu de choisir la route qu'elle a déjà traversé. Lors de l'achèvement d'un cercle de route, elle produit une quantité correspondante de la concentration de phéromone, sur la base de la distance totale de la route, et met à jour la concentration de phéromone sur le trajet passé.

Avec $\tau_{ab}(t)$ représente la concentration de phéromone sur l'arc (a, b) à l'instant t, la concentration de phéromone sur cette route à l'instant t + 1 est:

$$\tau_{ab}(t+1) = \rho \cdot \tau_{ab}(t) + \sum_{k=1}^m \Delta \tau_{ab}^k \quad (2)$$

où :

ρ - Une constante qui indique le facteur de quantité de phéromone, sa valeur est comprise entre 0 et 1, et $(1 - \rho)$ représente le coefficient d'évaporation de phéromone.

$\Delta \tau_{ab}^k$ - Indique la quantité de phéromone ajoutée sur la route (a, b) par la fourmi k entre l'instant t et t + 1.

$$\Delta \tau_{ab}^k = \begin{cases} \frac{Q}{d_{ab}} & \text{Si la fourmi } k\text{-ième passe par (a, b)} \\ 0 & \text{Sinon} \end{cases} \quad (3)$$

Q est une constante qui est la quantité totale de phéromone produite par chaque fourmi ; d_{ab} est la distance entre le nœud a et le nœud b.

Eq. (3) indique que l'accroissement de la concentration de phéromone est lié au coût de la route traversée.

2) Principe de sélection de Waypoints

La règle de transition consiste à orienter la direction de recherche de chaque fourmi vers la meilleure solution. Pour la fourmi k localisée au nœud a, la probabilité de choix de l'arc sortant (a, b) à l'instant t est définie par :

$$P_{ab}^k(t) = \begin{cases} \frac{\tau_{ab}^\alpha(t) \cdot \eta_{ab}^\beta}{\sum_{b \in N_{k(a)}} \tau_{ab}^\alpha(t) \cdot \eta_{ab}^\beta} & , \text{ si } b \in N_{k(a)} \\ 0 & \text{sinon} \end{cases} \quad (4)$$

où :

$N_{k(a)}$ - Ensemble de nœuds candidats qui peuvent être visités par la fourmi k située au sommet a.

η_{ab} - valeur heuristique, dite visibilité, définie par l'inverse du coût sur l'arc (a, b) ; $\eta_{ab} = 1/d_{ab}$.

α, β : Les deux principaux paramètres, qui contrôlent l'importance relative de l'intensité et de la visibilité d'une arête.

$\tau_{ab}(t)$ - Quantité de phéromones sur l'arc (a, b) à l'instant t.

Cette règle de transition donne la probabilité de choix du prochain nœud à sélectionner selon la quantité de phéromones et la valeur heuristique locale. Les paramètres α et β contrôlent l'importance relative de ces deux composants. Si $\alpha=0$ les fourmis choisissent le nœud ayant la meilleure valeur heuristique. En revanche, si $\beta=0$, le choix de probabilité ne dépend que la quantité de phéromones. La piste de phéromones est renforcée immédiatement dès qu'une fourmi complète une route. La quantité de phéromones ajoutée dépend de la qualité de route traversée.

3) Procédure de calcul

Les étapes d'utilisation de l'ACF pour déterminer la route optimale sont énumérées ci-dessous:

1. Initialisation de la quantité de phéromone sur tous les nœuds situés dans la zone de navigation, pour former la matrice initiale.
2. m fourmis sont prêtes à démarrer au point de départ A.
3. Chaque fourmi choisit le nœud suivant dans la matrice sur la base de la règle de transition d'état de l'équation (4), et atteint la destination à la fin, formant une route possible.

4. Calculer la fonction objectif d'une route possible de chaque fourmi sur la base de l'équation (1), et garder la route optimale.

5. Selon la fonction objectif et le principe de réglage de la quantité de l'équation (2), ajuster la quantité de phéromone de chaque point.

6. Vérifier si les conditions d'arrêt sont remplies (si le nombre de cycles établis ou la valeur minimum de la fonction objectif est atteint). Si les conditions sont remplies, la recherche est achevée, sinon, répéter la procédure à partir de l'étape (2) jusqu'à ce que les conditions d'arrêt soient remplies.

4) Calcul du coût de navigation

Le calcul des coûts de navigation doit tenir en compte des conditions environnementales, les zones maritimes dangereuses et les performances hydrodynamiques des navires.

a) *Comportement hydrodynamique du navire* : Les navires qui naviguent dans la Méditerranée sont affectés par le vent et les vagues. En conséquence, la vitesse réelle du navire dans les vagues sera certainement inférieure à celle dans l'eau calme. La dégradation de vitesse est la plus importante parmi les divers facteurs impliqués dans le calcul de route maritime [8]. Elle est présente dans de nombreuses formules de calcul de route, elle a une influence notable sur la précision de positionnement des navires. Une enquête menée par le Comité régional des pêches maritimes et des élevages marins de Bretagne en 2007 montre également une prise de conscience très claire de la part des professionnels de la navigation, de la nécessité de mettre en place un certain nombre de « bons gestes » pour faire des économies de carburant.

La formule retenue pour calculer le ralentissement de vitesse peut être utilisée pour produire des courbes de performance des navires ou des courbes de dégradation de vitesse, comme le montre la figure 2 ci-dessous. L'effet de la direction des vagues sur la dégradation de vitesse est différent en fonction de cette direction si elle est en face ou en arrière. La relation entre l'orientation du navire et les directions des vagues sont montrées dans la figure 3.

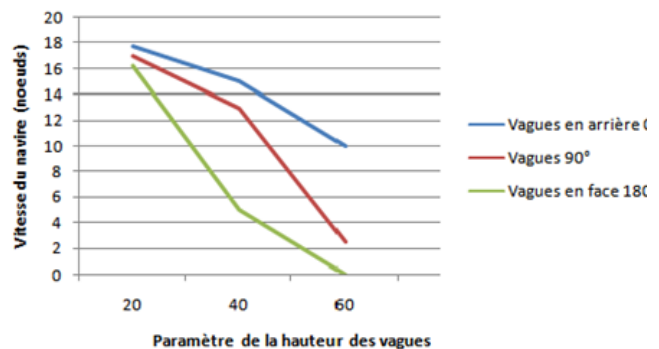


Figure 2 : Courbes de performance du navire

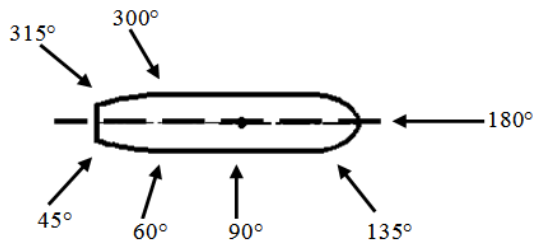


Figure 3 : Orientation du navire et la direction des vagues

Les principaux facteurs affectant la vitesse du navire sont le vent et les vagues, alors que la hauteur des vagues est étroitement liée à la vitesse du vent. Pour des raisons de simplicité, la formule empirique recommandée par l'Institut central de la recherche maritime russe pour calculer le ralentissement de vitesse du navire a été retenue. Les principaux facteurs inclus sont la hauteur des vagues, direction des vagues, et coefficient de performance du navire. La formule est exprimée comme suit :

$$V = V_0 \cdot (0.745h - 0.257q \cdot h) \cdot (1.0 - 1.35 \cdot 10^{-6} D \cdot V_0) \quad (5)$$

où :

V – la vitesse réelle dans la mer.

V_0 – la vitesse dans l'eau calme.

h – la hauteur des vagues.

q – l'angle entre le cap du navire et la direction des vagues.

D – la charge réelle du navire.

b) *Traitement des données de l'environnement* : Les données environnementales comprennent des données géographiques, de navigation, et des données météorologiques, c'est à dire les informations de base sur les zones de navigation pour le calcul des coûts. En combinant des calculs, des mesures de la distance de navigation et des données de l'environnement, différentes données de coût peuvent être obtenues. Le processus de traitement et de calcul est le suivant :

Les données géographiques : les frontières des terres et les limites de la mer et les zones de danger (par exemple, les plates-formes pétrolières ou les zones maritimes peu profondes) sont stockées sous forme de données vectorielles de type polygone dans le SIG. Nous avons initialisé les nœuds dans les zones terrestres et les zones des plates-formes pétrolières avec la valeur de 0, tandis que les autres nœuds sont définis par une valeur constante. Ainsi, les fournis ne passent pas par les nœuds dans la zone de terres et ne recherchent que dans la zone maritime.

Les données météorologiques : avec les progrès dans les nouvelles technologies de l'information et communication, de nombreuses informations météo sont alors facilement accessibles. Les navires en haute mer ne peuvent être connectés à l'internet que par des satellites, ce qui est relativement coûteux, il n'y a que certains navires qui utilisent cette méthode pour recueillir des informations météo. Les données GRIB (General Regularly-distributed Information in Binaryform) téléchargés de l'internet sont converties en un format SIG dans le système et utilisés pour l'optimisation de la route.

Les données hydro-météorologiques GRIB sont standardisées par l'Organisation mondiale des météo, et de nombreuses stations aux États-Unis et en l'Europe publient

des données météorologiques dans ce format. En général, les prévisions pour les 5 prochains jours sont disponibles, et ils sont mis à jour toutes les 6, 12 ou 24 heures. La résolution spatiale peut atteindre $0,5^\circ \times 0,5^\circ$. Ce format offre une précision pour la vitesse du vent, direction du vent, la pression atmosphérique, la température de l'eau de mer, la hauteur des vagues, la direction des vagues, et certaines autres informations météorologiques peuvent être incluses de manière sélective. Ces informations répondent adéquatement aux exigences du routage météorologique. L'information peut être téléchargée directement à partir d'internet à l'ordinateur de bord pour un traitement ultérieur, pour gagner du temps et améliorer la précision. Le ralentissement de la vitesse du navire est principalement causé par les vagues, la direction des vagues et la hauteur des vagues sont extraites des données GRIB pour le calcul de la route. L'effet du vent a été ignoré parce que sa relation avec les vagues est linéaire.

c) *Méthode de calcul* : la grille de données « la direction des vagues et la hauteur des vagues » est extraite du GRIB. Les données sont ensuite converties en cellules de la couche de données dans le SIG pour faciliter l'intégration avec les fonctions de calcul de SIG. Ensuite, l'utilisation des fonctions d'analyse spatiale SIG est faite pour calculer la distance dans toutes les cellules pour chaque route candidate, et de trouver les données de la hauteur des vagues et la direction des vagues des cellules passées. En utilisant la formule de ralentissement de vitesse du navire, la valeur de la vitesse perdue peut être calculée. Enfin, sur la base de la distance à travers les cellules, le coût en termes de temps ou en termes de consommation du carburant peut être calculé. Si le temps de navigation est considéré comme le principal coût, la formule de calcul est la suivante :

$$t_j = \frac{d_j}{v - w_j} \quad (6)$$

$$T_c = \sum_{j \in J} (t_j + d_j \times F_j) \quad (7)$$

où :

t_j - le temps requis pour passer la $j^{\text{ème}}$ cellule.

d_j - la distance de la route candidate pour passer la cellule j .

v - vitesse normale sans perturbation météo.

w_j - la vitesse perdue par les effets des vagues dans la cellule j .

T_c - coût total de navigation.

F_j - facteur de coût de distance additionnel dans la $j^{\text{ème}}$ cellule.

5) Contribution à l'amélioration des performances de calcul

a) *Limites de déviation de la direction maximale et de la longueur maximale de segment de route* : compte tenu du grand nombre de nœuds, beaucoup de capacités de calcul seront perdues. L'avantage de limitation de déviation de direction maximale et limitation de la longueur maximale de segment de route est d'éviter beaucoup de calculs inutiles. Nous avons limité l'angle de déviation de route à $\pm 45^\circ$ par rapport à la direction originale et la longueur maximale du segment de route ne doit pas dépasser celle qui peut être parcourue en un jour de navigation avec une vitesse courante.

b) *Détermination de la vitesse critique* : tout en effectuant des calculs de routage, on doit tenir compte de la vitesse critique du navire (vitesse maximale autorisée) lors de la navigation dans des conditions météorologiques différentes. Ainsi, le résultat de calcul peut produire la route optimale en respectant le principe de la sécurité. Le calcul de la vitesse critique est principalement lié à la hauteur des vagues et la direction des vagues. La formule suivante [9] est utilisée pour imposer des limites de la vitesse maximale sur la vitesse critique d'un navire pour assurer la sécurité de la navigation.

$$V_{\text{Limit}} = e^{0.13[\mu(q)-h]^{1.6}} + r(q) \quad (8)$$

où :

$$\mu(q) = 12.0 + 1.4 \times 10^{-4} q^{2.3}$$

$$r(q) = 7.0 + 4.0 \times 10^{-4} q^{2.3}$$

q – représente la direction des vagues.

h – représente la hauteur des vagues.

c) *L'ajout d'opérateurs de croisement pour l'algorithme génétique* : nous avons utilisé l'opération de croisement pour la recherche de la route optimale et d'autres routes d'une façon aléatoire à chaque itération. Si les deux fourmis passent par le même point (Waypoint) de la grille, alors une opération de croisement est réalisée systématiquement entre les deux routes (sauf les points de départ et d'arrivée) sur la base du même point de la grille, en combinant la première moitié de la route A avec la seconde moitié de la route B pour former une nouvelle route, et également la deuxième moitié de la route A avec la première moitié de la route B pour former une nouvelle route. Si les deux fourmis ne passent pas le même point de la grille, alors les opérations de croisement ne sont pas effectuées. Si, après l'opération de croisement, la route obtenue est meilleure que la route optimale actuelle alors l'information de la route optimale est mise à jour. L'ajout des opérations de croisement augmente la capacité de la fourmi pour trouver une meilleure route dans la recherche, ainsi que la variété des solutions.

d) *L'ajout d'opérateurs de mutation* : afin d'effectuer une exploration plus large de l'espace de recherche et d'éviter toute convergence prématurée ou disparition de la diversité des solutions, nous avons ajouté le concept d'opération de mutation à l'AG par la sélection d'un point de la grille de la route optimale actuelle et le muter avec un point aléatoire de la grille non utilisé, pour former une nouvelle route. Si la nouvelle route a un coût inférieur à celui de la route optimale d'origine, alors la route optimale d'origine est remplacée par cette nouvelle route.

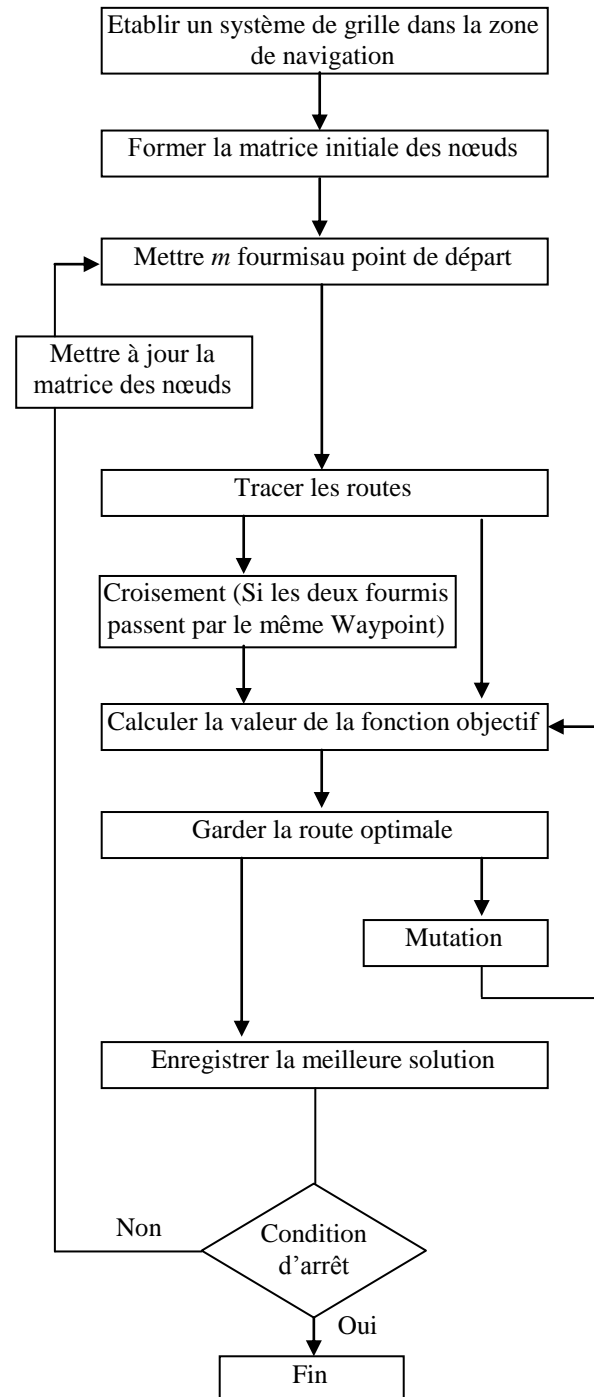


Figure 4 : Organigramme de l'approche hybride ACF et AG

IV. RESULTATS

Cette recherche expérimentale a simulé la route de navigation de Oran-Algérie à Bizerte-Tunisie, avec (35° 42' 54" N, 0° 37' 50" W) comme des coordonnées de point de départ, et (37° 16' 34" N, 9° 50' 44" E) comme des coordonnées de point final. Les données GRIB étaient utilisées pour la méditerranée en 17/03/2014. La vitesse de navigation est de 18 nœuds en mer calme. La route a été générée par la recherche de l'ACF ; elle a été générée en testant et en combinant les points de passage.

Les résultats simulés sont présentés sur la Fig. 5, où la route noire est la route orthodromique, la route rouge est la route

optimale produite par l'ACF, et les routes jaunes sont les routes générées dans le processus de recherche.

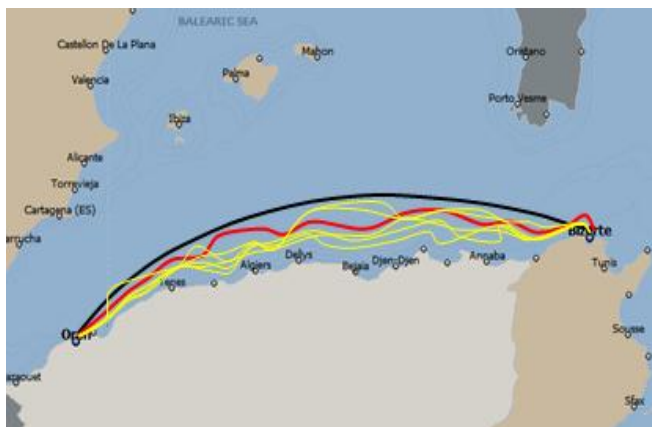


Figure 5 : Simulation de routes générées par l'Algorithme

Les routes maritimes ont été évaluées sur la base du temps de navigation total. A partir des expériences, il a été constaté que le choix approprié du nombre des fourmis et des paramètres peut améliorer l'efficacité de l'algorithme.

Lorsque le nombre de fourmis est faible, l'ACF ne montre pas de signes clairs de convergence, et la solution optimale est difficile à obtenir. Ceci est également vrai lorsque le nombre de fourmis est grand. Donc un gain de 45 minutes de navigation et un gain pour la consommation du carburant par rapport à la route orthodromique.

V. CONCLUSION

Dans ce papier, nous avons étudié le problème du routage optimal des navires en modélisant le déplacement et en dégageant une solution basée sur un algorithme d'optimisation qui optimise les déplacements effectués par les navires et qui tient compte à la fois de la sécurité et de l'économie. Pour cela, nous avons proposé un nouvel Algorithme de Colonie de Fourmis (ACF) avec l'intégration des concepts de l'Algorithme Génétique (AG). Le calcul et la recherche de l'information géographique sont fournis par le Système d'Information Géographique (SIG) développé.

Dans le futur, nous pouvons intégrer un mécanisme d'évitement de collision et non pas seulement les obstacles fixes (plates-formes pétrolières). Des recherches intéressantes Ito et al. [10], et Smierzchalski Michalewicz [3] et Tsou [11, 12] ont utilisé l'AG et ils ont obtenu de très bons résultats. Les environnements informatiques distribués et les applications de traitement parallèle peuvent encore améliorer l'efficacité de la résolution par le mécanisme de l'intelligence collective ACF.

REFERENCES

[1] J. Flecks, "Study Green Shipping, HVB Global Shipping", Hamburg, pp. 4-5, 2009.

nombre de fourmis est plus grand (20) et les paramètres α et β sont de faible valeur (≤ 3). Quand α et β sont plus grandes (≥ 4), la convergence est plus évidente, mais un grand nombre d'itérations est nécessaire. Lorsque le nombre de fourmis ≥ 50 et $\alpha, \beta > 4$, une meilleure convergence est atteinte, mais la solution est tombée dans un optimum local à de nombreuses reprises.

Il a été constaté que, lorsque le nombre de fourmis est relativement grand et les paramètres sont tels que $\alpha < \beta$ ($\alpha = 2, \beta = 3$), une meilleure convergence a été obtenue par l'ACF et une solution optimale proche a été plus facilement obtenue. Par conséquent, les paramètres pour dériver la route optimale sont définis comme suit:

$\alpha = 2, \beta = 3, \rho = 0.5, m$ (nombre de fourmis) = 100 et N (nombre de cycles) = 50.

Lors de la navigation en direction Est, en raison de l'influence des vents Nord-Est en provenance de l'Italie, la simulation ACF indique que la route de navigation devrait être déplacée au sud de la route orthodromique (Fig. 5). Si le navire a suivi la route orthodromique dont la distance est 561,4 miles nautiques, avec la vitesse en eau calme de 18 nœuds, le temps de navigation dans les conditions météorologiques était 32,5 h et la vitesse moyenne était de 17,26 nœuds. Malgré que la distance générée par l'ACF (égale à 568,3 MN) soit plus longue que la route orthodromique, le temps de navigation était seulement de 31,75 h et la vitesse moyenne était de 17,89 nœ

- [2] S. J. Bijlsma, "Minimal Time Route Computation for Ships with Pre-Specified Voyage Fuel Consumption", *The Journal of Navigation*, vol. 61, 2008, pp. 723-733.
- [3] R. Smierzchalski, Z. Michalewicz, "Modeling of ship trajectory in collision situations by an evolutionary algorithm," *IEEE Transactions On Evolutionary Computation*, Vol. 4, 2000, pp. 227-241.
- [4] M. Kefi Gazdar, "Optimisation Heuristiques Distribuée du Problème de Stockage de Conteneurs dans un Port," Thèse de doctorat, Ecole centrale de Lille ; Juin 2008.
- [5] S. Wei, P. Zhou, "Development of a 3D Dynamic Programming Method for Weather Routing," *International Journal on Marine Navigation and Safety of Sea Transportation*, Vol. 6, No. 1, 2012, pp. 79-83.
- [6] A. ZOGHLAMI, "Modélisation et conception de systèmes d'information géographique gérant l'imprécision," Thèse de doctorat, Université Paris 8 Ecole Doctorale Sciences Sociales, 2013.
- [7] R. Motte, S. Calvert, "On The selection of Discrete Grid Systems for On board Micro-based Weather Routing," *The Journal of Navigation*, Vol 43, No. 1, 1990, pp. 104-117.
- [8] R. Motte, R. S. Burns, and S. Calvert, "An Overview of Current Methods Used in Weather Routing," *The Journal of Navigation*, Vol. 41, No. 1, 1988, pp. 101-114.
- [9] F. Wang, C. Jia, "The Study on The Optimal Ship Routing," in *Chinese Journal of Dalian Maritime University*, Vol. 24, No. 2, 1998, pp. 61-64.
- [10] M. Ito, F. Zhang and N. Yoshida, "Collision avoidance of ship with genetic algorithm. Proceedings of 1999 IEEE International Conference on Control Applications, 1999, pp. 1791-1796.
- [11] Tsou, M.-C., Kao, S.-L., Su, C.-M.: Decision Support for Genetic Algorithms for Ship Collision Avoidance Route Planning. *The Journal of Navigation*, Vol. 63, 2010, pp. 167-182.
- [12] Tsou, M.-C.: Integration of a Geographic Information System and Evolutionary Computation for Automatic Routing in Coastal Navigation. *The Journal of Navigation*, Vol. 63, 2010, pp. 323-341.

TRACIS: a system for interpreting learners' traces

Amina ZEDADRA, Yacine LAFIFI, Safia BENDJEBAR

*LabSTIC laboratory, University 8 May 1945 Guelma
BP 401 Guelma 24000, Algeria*

zedadra_a@yahoo.fr, laf_yac@yahoo.fr, bendjebbar_s@hotmail.fr

Abstract— In Computer-Supported Collaborative Learning (CSCL) systems, the learners collaborate to reach the common goals. They left a lot of traces that concern all their pedagogical activities: learning, assessment, collaboration, tutoring, etc. The huge quantity of traces makes some tasks very difficult mainly the following-up of learners. The last is carried out by human actors called tutors. These actors found a lot of data on all the activities of their learners. So, how they can filter the pertinent traces? How can they interpret these traces in order to better help the learners? What are the tools to visualize these traces? The aim of this paper is to answer these questions by proposing a CSCL system that contains a new module of interpreting traces. This system is called TRACIS (TRACes Interpreting System). It is based on traces left by learners during their learning process. So, our main goal is to promote learning, simplifying the visualization and interpretation of these visualizations, and have a good use of traces made by the learners to ensure an effective monitoring in distance learning.

Keywords—Trace, CSCL, Interpretation, E-learning, Indicator, Collaboration.

I. INTRODUCTION

As it was mentioned by many authors and empirical studies, collaboration is an important activity in educational systems. This activity allows learners to share their knowledge and experiences. The main aim of this activity is to enhance learners' profiles (cognitive and behavioral). For doing this, many e-learning systems propose to their learners collaborative tasks that must be carried out in groups.

Computer-Supported Collaborative Learning (CSCL) systems provide learners with tools that facilitate the learning process. Learners can learn by collaborating in synchronous or asynchronous manner by using a lot of communication tools. Furthermore, in some CSCL systems collaborative scenarios are prepared by teachers. Learners can learn together learning objects' concepts, resolve collaborative exercises, carry out collaborative projects, prepare together common product, etc. With these different activities, the aim is to enhance the knowledge of each learner and improve his cognitive and behavioral profiles.

When using CSCL systems, the learners left a lot of traces. They concern all the learners' actions during their learning and collaboration process. Tutors can visualize learners' traces in order to assist them and to know their actions. In addition, these traces can be visualized by learners themselves. But, these traces are voluminous and concern many pedagogical activities (learning, assessment, tutoring, etc.). So, a step of interpreting of these traces is required.

The goal of this paper is to present a CSCL system that takes into account the process of interpreting traces. In other words, it can give more information about the actions carried out by the learners during any period. For doing this, a set of indicators are proposed that concern the main learners'

activities and traces. Furthermore, for facilitating the interpretation step both graphic and textual interpretation is provided by the interpreting system, which is called TRACIS (for TRACes Interpreting System).

In TRACIS, we focus on the tutor who has a leading role in monitoring learners and support the collaborative learning activities. It can be used by students and tutors at any higher education establishment that adopts the tutoring task (it is an essential task for students of the first cycle in Algerian universities). Furthermore, it has many features for facilitating the filtering of pertinent traces and their interpretation. In fact, the developed system provides each actor with tools of perception and visualization of traces to adapt them to his role and his tutoring style. The perception of what happens during the learning activity plays a crucial role in enabling learners to make decisions they consider suitable and carry out their tasks.

TRACIS is developed at LabSTIC laboratory (<http://www.labstic.com/>). We present some screenshots about the developed system and the appreciations of some students who used the system. Actually, the system is under test at a higher education establishment.

The rest of the paper is organized as follows. In the section 2, we present some research works about the interpretation of traces. Section 3 presents the main architecture of TRACIS system. In section 4, we present some screenshots about the developed system. Section 5 is reserved to the conclusion and the future work.

II. RELATED WORKS

There are many definitions associated to the term of trace. According to (Champin and Prié, 2002), traces are "a sequence of states and transitions representing the user activity". (Choquet and Iksal, 2007) define the trace in the field of the re-engineering of the human learning environment as "a teaching object as well as the pedagogical resources or scenarios". In another term, Jermann and his colleagues defined a trace as (Jermann et al., 2001) "an observation or a recording of the interaction of learners with a system for an analysis".

Also, the term of interpretation is defined in different ways according to the context of its application. As result, it has various definitions. (Daele and Berthiaume, 2011) defined the interpretation as "an analysis of the learners' production and an attribution of a value to the learning traces that they provided". According to (Lund and Mille, 2009), "the term of interpretation is to provide a formal semantics to a symbolic description".

There are some works that aim at interpreting traces. (Cheype, 2006) interprets the traces of learners in order to detect if the trace is a succession of actions which tend to lead to the success or the failure of an exercise. This approach is used to calculate indicators that are calculated by

using data mining techniques. Heraud and his colleagues (Heraud et al., 2005) propose to understand the learners' behavior in order to improve their pedagogical scenarios by an interpretation of their activities traces (use of the courses or the forum during the resolution of a set of exercises). GISMO (Mazza and Milani, 2005) is a tool, which traces and visualizes the learner's activities. It uses the information to make graphical representations. The latter are handled by the instructors to know what happens in the distance classes. May and his colleagues (May et al., 2009) propose a visualization approach to offer to the users a simple sight but representative of their activities. The system expresses the learners' traces to help the participants in their learning processes. France and his colleagues (France et al., 2007) present a traces' visualization interface which allows the tutor to observe and adapt the learners group' activity within a human learning environment called ClassroomVis. The latter offers a double interest: control the learning situation and the dynamic interaction between learners.

TABLE I. Summary of the related works.

Authors	Visualization (yes/No)	Visualization for?	Interpretation (Yes/No)	Interpretation for?
[Mazza and Milani, 2005]	Yes	Instructors	Yes	Instructors
[France et al., 2007]	Yes	Tutors	Yes	Tutors
[May et al., 2009]	Yes	Users	No	/

[Cheype, 2006]	Yes	Tutors	No	/
[Heraud et al., 2005]	Yes	Designers	No	/

III. ARCHITECTURE OF THE DEVELOPED SYSTEM

The developed system (i.e. TRACIS) is a CSCL system. So, it has the main tools of e-learning systems and those facilitating the collaboration process. Its main part is the traces interpreting sub-system, which uses some interpretation rules in order to provide the teachers and the tutors with more information about the learning paths of each learner. These rules can be updated any time by the tutors or the administrator of the system.

Figure 1 presents the main architecture of the system (TRACIS). It has four interfaces that are associated to the human actors of the system: administrator, teacher, tutor and learner. In addition, it holds a data base that has all the information about the human actors, their traces, the interpretation rules, the learning objects, the assessment results, the collaboration contents, etc.

TRACIS possessed three sub-systems:

Learning Management System (LMS) that manages all the tasks about the learning and the assessment processes.

Collaboration Management System (CMS) that manages all the tasks related to the collaboration activity and the tools facilitating the communication among learners.

Traces Interpretation System (TIS) that takes into account the task of interpreting the traces left by learners.

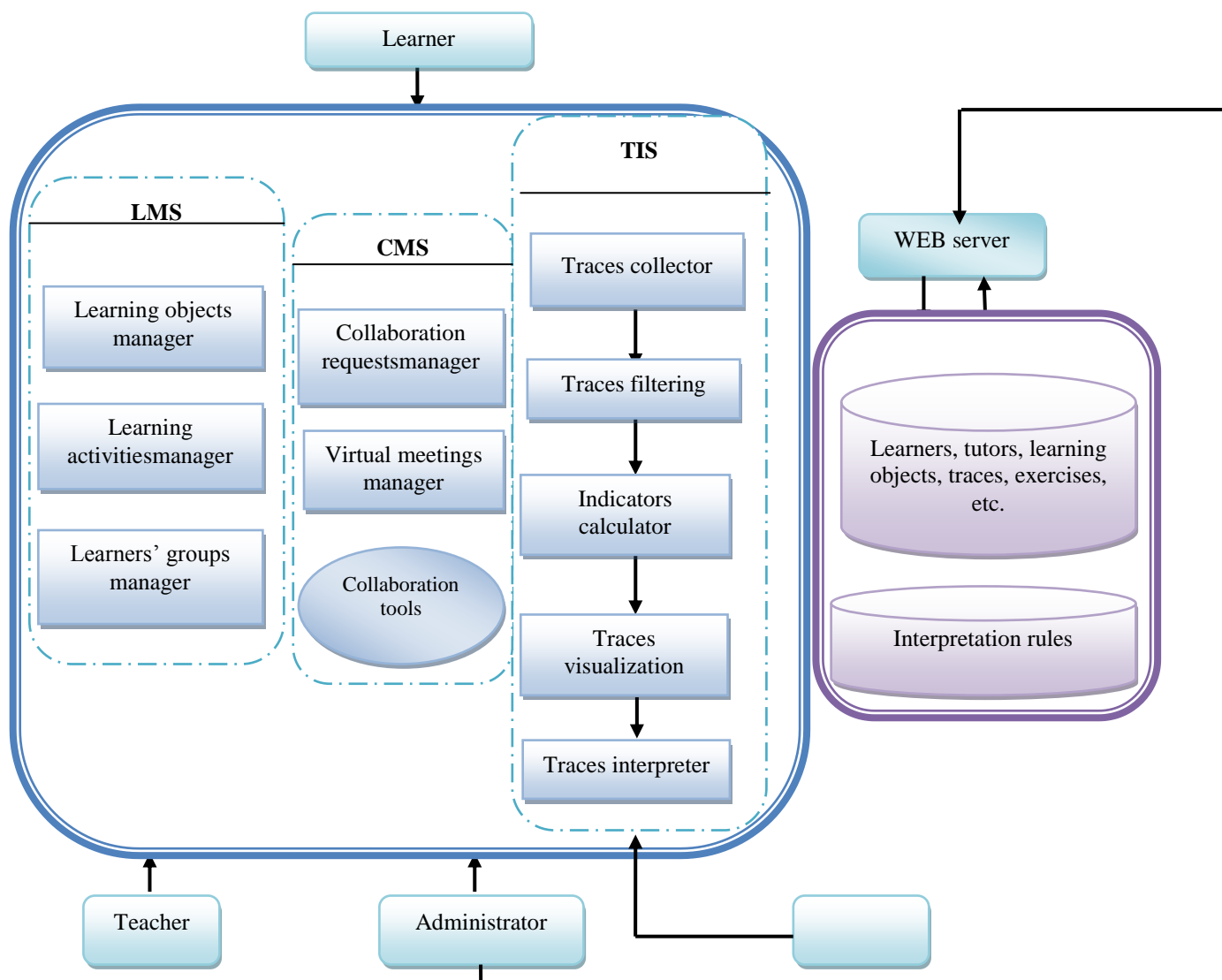


Fig. 1. System Architecture.

A. Learning Management System (LMS)

This part has three modules. The first one is the learning objects manager, which is responsible of managing the learning objects (insertion, suppression, search, etc.). The second module is the manager of learners' groups, which takes into account the process of the organization of learners into groups. Several methods are adopted by TRACIS for forming groups: randomly, formation by teachers or according learners' choices. The third and the last module is the manager of learning activities. This module provides the authors (teachers) with tools that facilitate the management of the collaborative activities. The latter are proposed to enhance the learning quality of learners by collaborating.

B. Collaboration Management System (CMS)

CSCL systems are designed to support the collaboration among learners. TRACIS has a subsystem that takes into account the collaboration process. So, it can manage the collaboration requests sent by the learners and the virtual meetings. Each learner can demand collaboration with another learner by specifying the object of collaboration. All the information on the collaboration requests are saved in the data base for a future use. They are used for calculating the

behavioral profiles of learner. In fact, the system recorded the main information about each demand of collaboration (sender, receiver, date, preferred tool of collaboration, object, and the appreciation of learners in the case when a collaboration is carried out).

C. Traces Interpretation System (TIS)

This sub-system has five modules. The first one is the *collector of traces*, which gathers all the traces left by the learners when using TRACIS. These traces concern: the learning activities, the assessment exercises, the collaboration requests, the use of research engine, etc. In CSCL systems, these traces are voluminous. So, a filtering step is necessary. This action is carried out by the second module, which is the *traces filter module*. It aims at filtering the traces and let in the data base only the relevant ones.

In order to interpret the traces, some indicators are calculated. This task is carried out by the *indicators calculator module*. The latter is responsible of calculating all the indicators that will be used for facilitating the interpretation process. In fact, the interpretation rules are the values of these indicators for knowing the best interpretation of traces.

The traces can be visualized by a module called *traces visualization module*. It provides the users with some visualization tools (graphical, textual, or hybrid).

Finally, according to the values of the indicators, the *tracesinterpreter* carried out its tasks. It is responsible of the main task of TRACIS, which is the interpretation of learners' traces.

Interpretation rules:

In our system, the tutors have many roles: pedagogue, guide, organizer, etc. By observing the learners' traces, the tutors can send messages containing advice to learners according to some interpreting rules. These rules used the values of some indicators that are related to the different pedagogical activities of the learners. We give some of the indicators used by TRACIS.

EXAMPLES OF INDICATORS.

N	Indicator	Pedagogical Activity
1	Number of daily collaboration requests.	Collaboration
2	Number of daily received collaboration requests.	Collaboration
3	Period of each collaboration process.	Collaboration
4	Number of emails sent by the learner.	Communication/ Collaboration
5	Number of emails received by the learner.	Communication/ Collaboration
6	Number of solved collaborative exercises.	Collaboration/Assessment
7	Percentage of responses added to the forum.	Communication/ Collaboration
8	Percentage of time spent by learner in consulting learning objects.	Learning

9	Percentage of learner's presence in the system.	Learning
10	Percentage of requesting assistance to other members of the groups.	Tutoring
11	Percentage of the time spent by learner in resolving assessment exercises	Assessment

In order to better explain our contribution, we give some examples of rules used to send messages having advice to learners.

If the number of solved collaborative exercises in a week < 2 then send a message ('It is better to solve more exercises with your teammates').

If the percentage of the time spent on one learning object compared to the total period of the session $< 50\%$ then send a message ('It is better to consult more learning objects').

If the number of access days to the system per week < 3 then send a message ('Visit the system and benefit from its good features').

IV. IMPLEMENTATION OF THE DEVELOPED SYSTEM

TRACIS was implemented at LabSTIC laboratory at Guelma University. It is under evaluation by license students of Computer science department (specialty: "information systems"). Each learner can get access to the system from any computer connected to the intranet of the university. He must be enrolled successfully before beginning the use of the system. The tutors can see all the traces of their learners. Furthermore, they can see the interpretation of these traces. Figure 2 presents a screenshot taken from TRACIS. It presents the interpretation of learning traces. As it is mentioned by the figure, the concerned learner didn't access to the learning space from May 12th until May 18th. Between May 18th and June 4th, he accessed the system many times.

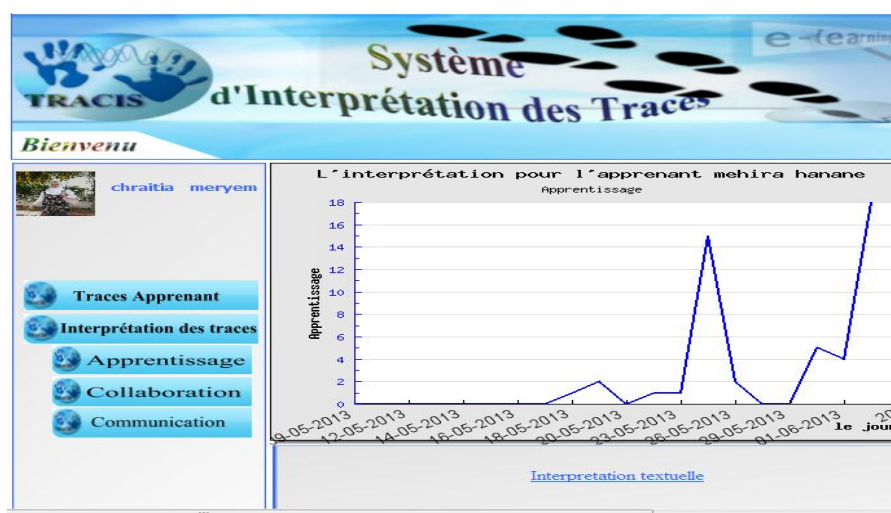


Fig. 2. Graphical and textual interpretation of learning traces.

Figure 3 presents graphical and textual interpretation of collaboration traces. This form of visualization helps the

tutors to take the right decision about the message content that will be sent to the concerned learner.

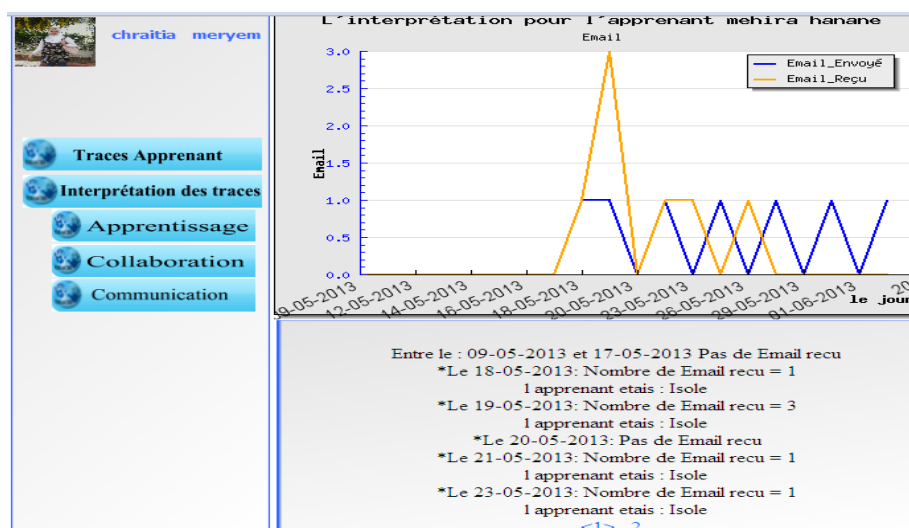


Fig. 3. Collaboration traces interpretation.

The system is under evaluation in order to extract the effectiveness of the interpretation process and the rules used in the actual version. The appreciations of some tutors who used the system are very encouraging. Furthermore, there are some problems cited by the tutors. The encountered difficulties will be corrected in the future version.

V. CONCLUSION AND FUTURE WORKS

In this research, we have proposed an interpreting subsystem that can be used in a collaborative learning system. It aims at facilitating the tutors' tasks and assisting learners during their pedagogical activities (learning, assessment and collaboration). Furthermore, it provides tutors with both textual and graphical interpretation of learners' traces according to some interpreting rules that can be updated by the tutors themselves. These rules used a set of indicators that are calculated from the traces left by learners.

The developed system (i.e. TRACIS) holds the main features of computer supported collaborative learning systems. As it is mentioned in its architecture, it has some modules that take into account the groups formation, the learning objects management (creation, downloading, updating, etc.), the communication tools management, the knowledge assessment, the collaborative activities management, etc.

TRACIS is under evaluation at a higher education institution. The first results shows that it can help tutors in their tasks by providing them with only the relevant traces among all the traces left by the learners. In addition, it presents to them an interpretation of these relevant traces. This task become automatic and let the tutors concentrate on their main tasks (i.e. tutoring) instead of interpreting manually the actions of their learners. This novel feature is our main contribution in this paper. In fact, the system interprets automatically all the relevant traces of the learners according to a set of interpretation rules which used some parameters. Actually, the system uses some parameters that were definite in collaboration with some researchers at psychology department of Guelma University.

As future work, we propose to conduct an experiment with university students and tutors in order to extract the

advantages, the limits and the appreciation of both tutors and students. Also, we propose to conceive a technique for automatic filtering traces and conceive a language for interpreting traces independent of their use. Finally, we propose to conceive other forms of traces representation and other interpretation methods based on the semantics of these traces.

VI. REFERENCES

- Champin, P.A. Prié, Y. (2002). Musette : un modèle pour réutiliser l'expérience sur le web sémantique. Journées scientifiques Web sémantique, October 2002, Paris, France.
- Cheype, A. (2006). Recherche de motifs séquentiels pour guider l'interprétation des traces d'apprentissage dans un EIAH. 1ere Rencontre Jeunes Chercheurs en EIAH, RJC-EIAH 2006, Mars 11-12, Every, France, pp. 123-130.
- Choquet, C. Iksal, S. (2007). Modeling tracks for the model driven reengineering of a TEL system. Journal of Interactive Learning Research (JILR), Vol. 18, pp. 161-184.
- Cram, D., Jouvin, D., Mille, A. (2007). Visualisation interactive de traces et réflexivité: application à l'EIAH collaboratif synchrone EMEDIATHEQUE. Sciences et Technologies de l'Information et de la Communication pour l'Enseignement et la Formation, Special issue: Analyses des traces d'utilisation dans les EIAH, Vol. 14. Available on « <http://sticef.univ-lemans.fr/num/vol2007/06-cram/sticef2007cram06.htm> ».
- France, L., Heraud, J.M., Marty, J.C., Carron, T. (2007). Visualisation et régulation de l'activité des apprenants dans un EIAH tracé. Environnements Informatiques pour l'Apprentissage Humain EIAH'2007, June 27-29, Lausanne, Suisse, pp. 27-29.
- Jermann P., Soller, A., Muehlenbrock, M. (2001). From mirroring to guiding: A review state of the art technology for supporting collaborative learning. Proceedings of the First European Conference on Computer-Supported Collaborative Learning.
- Heraud, J.M., Marty, J.C., France, L., Carron, T. (2005). Une aide à l'interprétation de traces : application à l'amélioration de scénarios pédagogiques. Environnements

Informatiques pour l'Apprentissage Humain EIAH'2005, March 25-27, Montpellier, France.

May, M., George, S., Prévôt, P. (2009). Tracer, analyser et visualiser les activités de communications médiatisées des apprenants. JOCAIR 2008, Journées Communication et Apprentissage Instrumentés en Réseau, August 27-28-29, University of Picardie Jules Verne, Amiens, France, pp. 27-29.

Mazza, R., Milani, C. (2005). Exploring usage analysis in learning systems: Gaining insights from visualisations. The Workshop on Usage analysis in learning systems, the twelfth International Conference on Artificial Intelligence in Education, July 18-22, Amsterdam, The Netherlands, pp. 65-72.

Lund, K., Mille, A. (2009). Traces, traces d'interactions, traces d'apprentissage : définitions, modèles informatiques, structurations, traitements et usages. Analyse de traces et personnalisation des environnements informatiques pour l'apprentissage humain, éditeur Hermès Sciences Publications, pp. 21-66.

Daele, A., Berthiaume, D. (2011). Choisir ses stratégies d'évaluation. Research report, February 2011, Lausanne University, Suisse.

Towards Building Web Service Ontology:

A WSDL Reverse Engineering Approach

Houda EL BOUHISSI

Department of Computer Sciences
EEDIS Laboratory
Sidi-Bel-Abbes University, Algeria
Houda.elbouhissi@gmail.com

Mimoun MALKI

Department of Computer Sciences
EEDIS Laboratory
Sidi-Bel-Abbes University, Algeria
Mimoun.Malki@gmail.com

Abstract— The Semantic Web Services technology is used to describe the Web services' capabilities and content in a computer-interpretable language and improve the quality of existing tasks, including Web Services discovery, invocation, composition and monitoring. This paper addresses the topic of defining a knowledge based system for representing Semantic Web Service ontologies according to the WSMO conceptual model and proposes a software engineering approach to this problem using an existing Web Service. The proposal uses a reverse engineering technique and a similarity measure starting from a WSDL File of an existing Web Service till modeling WSMO Ontology specified in WSML language. The paper also presents a tool to support that technique throughout a real-life case study.

Index Terms—*Web Service, Semantic Web Service, Ontology, WSDL, Reverse engineering.*

I. INTRODUCTION

Current technologies allow usage of Web Services but only with syntactical information descriptions, therefore, it only provides a set of rigid services that cannot adapt to a changing environment without human intervention. A possible solution to all these problems is likely to be provided by converting Web Services into Semantic Web Services to be semantically marked-up software resources that can be published, discovered, composed and executed across the Web in a task driven semi-automatic way.

Today, organizations are increasingly forced to modernize using Semantic Web Services which, in recent years, become one of the most effective, efficient, and economical means to make intelligent systems.

This migration calls for reverse engineering of Web Services to Semantic Web services. However, there are few approaches that consider Services' ontologies as the target for reverse engineering. A majority of the work on reverse engineering has been done on tools that require a previous knowledge of the Web Service Application, however few tools use WSDL File as resource information. As an attempt to fill gap in this area, we propose a novel approach to reverse engineering of Web Services to Service's ontology.

This paper continues our previous research on domain knowledge driven Web Service Analysis [1] and describes a reverse engineering process for building Semantic Web Service Ontology according to

the Web Service Modeling Ontology conceptual model.

The proposed approach deals with a software engineering technique that consists of extracting useful information from a WSDL file of an existing Web Service in order to build Web Service Ontology specified in the Web Service Modeling language (WSML).

Our approach is based on the idea that semantics of a Web Service can be inferred, without an explicit analysis of the Web Service code. Rather, these semantics can be extracted by analyzing WSDL description File, which is the most popular document that describes a Web Service application.

The semantics are supplemented with the domain ontologies and user "head knowledge" to build a WSMO ontology. Our approach can be applied to migrating Web Service application, which is usually described by the WSDL File to the ontology based Semantic Web.

The remainder of this paper is structured as follows: Section 2 summarizes the Semantic Web Services related works. In section 3, the proposed approach will be described in detail. In Section 4, an experimental test of the proposed techniques is reported. Finally, section 5 concludes the paper and gives future directions of the on-going project.

II. BACKGROUND

Several approaches have already been suggested for adding Semantic to Web Services. Semantics can either be added to currently existing syntactic Web Service standards such as UDDI and WSDL, or services can be described using some ontology based description languages. The Major initiatives in the area of SWSs are documented by W3C member submissions, such as, OWL-S [2], WSMO [3] and WSDL-S [4].

Ontology Web Language for Services (OWL-S) is a description language that semantically describes Web Services using OWL ontologies. OWL-S services are mapped to WSDL operations, and inputs and outputs of OWL-S are mapped to WSDL messages.

The Web Services Description Language - Semantic (WSDL-S) augments the expressivity of WSDL with semantics such as domain ontology in an arbitrary semantic representation language.

The WSDL-S proposal was superseded by Semantic Annotations for WSDL (SAWSDL) [5] which is restricted and homogenized version of WSDL-S in which annotations like preconditions and effects have not been explicitly contemplated.

The Web Service Modeling Ontology (WSMO) provides ontological specifications for the description of Semantic Web Services. One of the main objectives of WSMO is to give a solution to application integration problems for Web Services by providing a conceptual framework and a formal language for semantically describing all relevant aspects of Web Services.

Furthermore, WSMO is the only standard for which there exist several implementation environments which aim to support the complete standard. For these reasons WSMO is used as our Semantic Web Services technology throughout the rest of this paper. We will explain next the concepts of the WSMO approach in detail.

III. THE PROPOSED APPROACH

The main contributions of the approach presented in this paper can be summarized as follows:

1. We propose a conceptual modeling approach for the specification of ontologies. The approach is based on WSDL File of an existing Web Service and semantic similarity measures using WordNet;
2. By using the proposed approach, we describe a set of Web based software tools allowing the developer to define WSMO ontologies and the final user to explore them.

A. The WSMO Framework

WSMO is a conceptual model for describing Semantic Web Services. It consists of four components, describing semantic aspects of Web Services: ontologies, Web Services, Goals, and Mediators. Each of these WSMO Top Level Elements can be described with non-functional properties, such as, creator, creation date, format, language, owner, rights, source, type, etc.

Web Services element in WSMO describes semantic aspects of the already defined Web Services. Also, goal element expresses the user requests. The correct invocation of the right Web Service is done by means of elements called Mediators. These elements resolve heterogeneities between the interacting parties both at data level by providing ontologies mapping, and at process level by resolving communication mismatches.

Although this paper is related to ontologies specification, it briefly describes all the WSMO elements.

Ontologies provide a formal and explicit specification of the application domain and of all the data used by the other components. Optionally, they may be described by non-functional properties and may import existing ontologies.

However, the WSMO ontology is composed of:

1. Concepts, describing the ontology domain possibly organized in a hierarchy,

2. Relations, representing further connections among concepts,
3. Instances of concepts and relations, setting values to their attributes and parameters respectively,
4. Axioms for further definition of concepts and relations through logical expressions.

Furthermore, WSMO comes along with a Web Service Modeling Language Modeling language (WSML¹³) and a Web Service eXecution Environment (WSMX¹⁴).

The WSML is a formalization of the WSMO ontology and providing a language within which the properties of Semantic Web Services can be described.

WSMX provides an architecture including discovery, mediation, selection, and invocation. In addition, it has been designed to include all the required supporting components enabling an exchange of messages between requesters and the providers of services.

B. Motivation

Among the Graphical User Interface tools for building and managing ontologies compliant to WSMO, we distinguish the WSMO-Studio [6] and the Web Services Modeling Toolkit (WSMT) [7].

However, WSMO Studio is a WSMO editor for building the aspects of Semantic Web Services, available as Eclipse-based plug-ins. Its main functionalities include definition of ontologies, Web Services, Goals and Mediators, and services composition through graphical user interfaces.

We focus our analysis in the ontologies editor since our interest is on the ontologies building. WSMO Studio stores ontologies in WSML, the representation language of WSMO.

The tool supports interaction with repositories for import/export ontologies, but the editing is done on local copies on the user's machine.

At the moment, no extensions are offered to allow concurrent access and editing to a shared repository. Whereas, the Web Services Modeling Toolkit (WSMT) is an Integrated Development Environment (IDE) for Semantic Web Services developed for the Eclipse framework. The WSMT aims at aiding the developers of Semantic Web Services through the WSMO paradigm by providing a seamless set of tools to improve their productivity. The IDE focuses on three main areas, namely engineering of WSMO descriptions, creation of mediation mappings and interfacing with Semantic Execution Environments (SEEs) and external systems.

These tools produce ontologies in a completely manual manner where the user participation is fundamental at all stages of creation. Both tools are based on a prior knowledge and comprehension of the user, that is, the process involves select choices or introduces useful information.

This process is costly and time-consuming. In addition, currently, there is no automatic or semi-automatic tool for creating WSMO ontologies.

¹³ <http://www.w3.org/Submission/WSML>

¹⁴ <http://www.wsmx.org>

For these reasons, this paper describes a proposed approach to create a semi-automatic tool that uses the information provided by the WSDL file to create WSMO ontologies.

C. Methodology

Traditional reverse engineering tools extract knowledge from source code and software documentation [8]. However, this approach is rather limiting as often information concerning how the code was developed and the rationale for its design are lacking. Moreover, a piece of source code may be cryptic due to a lack of developer comments. The proposed approach in this paper is to use the description of an existing Web Service (WSDL File) itself to specify ontology according to the WSMO conceptual model. Therefore, this proposal proceeds mainly in two principal stages (figure1):

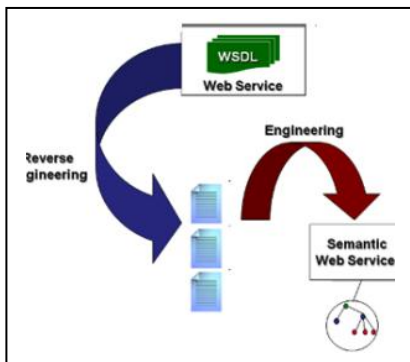


Fig. 1. Main stages of the proposal

- A stage of reverse engineering for the identification of the useful information in the WSDL File.
- A stage of engineering for the construction of the Web Service Ontology according to WSMO conceptual model

Our approach uses a WSDL File as input, and goes through five basic steps: (1) Entities identification to extract useful information from a WSDL File, (2) analyzing the extracted information by applying mapping rules to create the backbone of the ontology, (3) Semantic enhancement of the useful information using domain ontologies for applicability and consideration of the standardization problem, (4) Building ontology by translating the formal model into ontology element specified in the WSMO language, and (5) Validation of the produced element.

The proposed approach does not require that Web Service change at all, but adds a semantic layer to an existing Web Service. Therefore, it reduces the efforts and the cost to build a WSMO Ontology by reengineering without paying attention to the source code or how the application has been built.

The proposal is divided into the following five high-level phases of building the Semantic Web Service ontology process, see Figure 1:

a) **Phase 1: Entities identification.** This phase deals with the information required to create the WSMO ontology (concepts, attributes, relationships, and axioms) starting from an existing WSDL File.

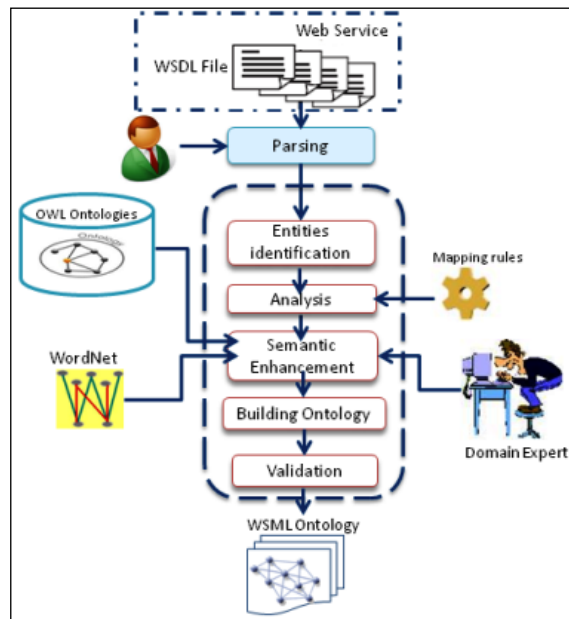


Fig. 2. System Architecture

In this Phase, we are interested to the XML Schema part of the WSDL File which is a description of data types of the Web Service's input and output. In this schema we can find the definition of ComplexTypes, SimpleTypes and the declaration of elements. However, their explanations are as the following:

- **Simple Type Definition:** XML Schema provides a wide range of built-in data types, for example, string, integer, Boolean, float, decimal, etc, these are examples of one form of simple type definitions. Another form of simple type definitions can be used to provide constraints on the values of built-in types. For example, it may be necessary to restrict the allowed values of the positiveInteger data type to a particular maximum value.
- **Complex Type Definition:** Can be used to: define a data type composed of sub-elements of other data types; define the allowed structure of child elements using the keywords all, sequence and choice; or extend or restrict the definition of an existing complex type. Additionally, the values of elements can be accompanied by constraints on their values.
- **Attribute Declarations:** Attributes can be either global or associated with a particular complex type definition. Attributes are an association between a name and a simple type definition.
- **Restrictions** (also called facets) are used to define acceptable values for XML elements or attributes. We use the restriction element to indicate the existing (base) type, and to identify the 'facets' that constrain the range of values.

The extraction phase is a full automatic process, we identify the information between the <element> and </element> according to Complextype, Simple type, and Attribute definition.

Also we enumerate all the restriction statement options. These information concern name, type, and restriction attribute. All the extracted information is stored in a XML file for further use.

For example, we assume that the XML schema part of the WSDL File contains the example definitions (see figure 3). We retrieve the following information: Complextype, Simpletype, shipto, name, string, address, string, city, string, country, string, restriction, enumerate Germany and Canada.

```
<xs:element name="shipto">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="name" type="xs:string"/>
      <xs:element name="address" type="xs:string"/>
      <xs:element name="city" type="xs:string"/>
      <xs:element name="country" type="xs:string"/>
    </xs:sequence>
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="Germany"/>
        <xs:enumeration value="Canada"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:complexType>
</xs:element>
```

Fig. 3. Example of XML schema part of a WSDL File

b) Phase 2: Analysis Phase. This step focuses on the mapping of retrieved information in the previous step using mapping rules. The mapping engine deals with a set of transformation rules and analyzes the information. This phase produces a set of concepts and axioms. Definition of each concept is accompanied by his sub concept, its components with their attributes name and type. The mapping produced is roughly based on the following rules (see Figure 4):

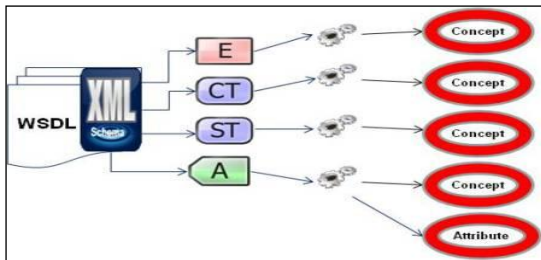


Fig. 4. Mapping Rules Definition

• **Rule 01: Simple type definition.** If a simple type is used to create a new type based, we create a new concept with the same built-in type (see figure 5 for an example. Also, if the simple type participates in the definition of a complex type, it will be mapped to a property (attribute) of the complex type with the main built-in type.

```
<xsd:element name="age">
  <xsd:simpleType>
    <xsd:restriction base="xsd:positiveInteger">
      <xsd:maxExclusive value="35">
    </xsd:restriction>
  </xsd:simpleType>
</xsd:element>
```

Fig. 5. Example for the Rule 01

• **Rule 02: Complex type definition.** Complex type definitions can obtain sub-components that are a mixture of simple elements, attributes and other complex type definitions. We propose to map each complex type to a concept in the WSMO Ontology. Sub-components with simple type built-in are mapped to attributes with the same built-in type and attributes are mapped to attributes with the same built-in type. If the sub-component is itself a complex type, here we proceed in a recursive manner, we create first the corresponding concept, and then the sub-components are mapped to attributes with the built-in type (see figure 6) which contains a definition assumed to be within the XML Schema part of the WSDL File. Finally, the complex type embedded in another complex type is mapped at one hand to sub-concept of the complex type and at the other hand to a concept.

```
<xs:element name="Customer">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="Dob" type="xs:date">
      <xs:element name="Address" type="xs:string" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Fig. 6. Example for the Rule 02

• **Rule 3: Attributes.** An attribute may be associated directly to the root or embedded in a simple or complex type. If an attribute depends to the root, we propose to create a new concept with the built-in type. If the attribute is embedded in a simple or a complex type, it is mapped to an attribute of the concept of the complex or the simple type.

• **Rule 4: Restriction element for data type.** Each restriction is mapped to an axiom, with the corresponding option. As an Example of restriction definition, in figure 3, the CountryConstraint which defines the corresponding value of the Country.

The results of the analysis phase are stored in a XML file which is mainly designed by the tags <complexType> . . . </complexType> and <simpleType> . . . </simpleType>. This formalization facilitates better the translation into the WSMO language.

c) Phase 3: Semantic Enhancement. In the process of identifying entities in the document, it is possible that we find values for attributes or relationships that were not previously present in the knowledge base. Enhancing the existing metadata could be as simple as entering values for attributes, in which case they could be automated; or as complex as modifying the underlying schema, in which case some user involvement might be required.

The third step of our proposed is calculation of semantic similarity measure which aims to quantify how much two

concepts from the XML file produced before and domain ontology are similar.

The similarity analysis of these concepts is used through WordNet dictionary which will give us a standardized and complete synonym set and classifies the entities. We use WordNet¹⁵ based similarity measures [9] such as Path, Resnik, Lin and Jiang. WordNet is an online lexical database designed for use in a program.

The use of ontology allows greater expressiveness and provided rich semantic concepts. Measuring semantic similarity between words has been playing an important role in many research areas such as Artificial Intelligence, Linguistics, Cognitive Science, and Knowledge Engineering. Measuring the semantic similarity or distance between words is a process of quantifying the relatedness between the words utilizing the knowledge obtained from certain information sources.

For this reason, we use the WordNet as a taxonomic reference, so the idea is to compare the concept of the XML file already created with domain ontology, using the WordNet. We consider that to choose candidates concepts is enough to resolve semantic interoperability. Figure 7 depicts the algorithm of the semantic enhancement process. We have an XML file corresponding to the data types of the Web Service expressed, in concept, attribute, restriction and type. We are interested to only the concepts that have not sub concept, as concept candidates because we assume that is enough to identify if the produced WSMO ontology is well defined.

We also have an uploaded an ontology containing concepts according to any domain and each concept may possibly have a set of attributes. Subsequently, we perform similarity calculations to identify the semantic concepts of the ontology. We define a threshold value for prior semantic similarity measure. The threshold is a value between 0 and 1, the value 1 indicating that the two entities are completely similar.

Algorithm 1: Semantic enhancement

Input: XML File (i Candidates concepts)
 Concepts of Domain Ontology
 (j Candidates concepts)
 A Similarity measure
 Threshold value for the similarity measure
Output: List of concepts
Begin
 Create a vector containing the concepts of domain ontology (j concepts)
For each concept C_k of the XML File
 (k=1 to i do) **do**
For each concept CO_m m=1 of the domain
 ontology (m=1 to j) **do** Calculates the distance between
 C_k and CO_m
If (the similarity = the threshold) **then** supersedes the
 concept c_k by the concept CO_m
EndIf
EndFor
EndFor
End.

Fig. 7. Algorithm 1 : Semantic enhancement

The threshold value is an important decision point. If it is very low, between [0.0, 0.5], many concepts can be wrongly matched i.e. false positives can be returned by function similarity (c1, c2). On the other hand, if this value is high, between [0.7, 1.0], many concepts which might be matched are not caught, i.e. many false negatives can be discarded.

Second, we must choose a method to compute the semantic similarity measure as described above. If the semantic similarity measure between the concept of the XML file and the ontology concept is greater than or equal the threshold, we can consider that the concept belongs to the domain of the ontology concept, therefore this concept can be superseded by the concept of the of the domain ontology and at last its sub concept is retrieved for adding in the XML file.

d) Phase 4: Building Ontology. Ontologies and Semantic Web Services need formal languages for their specification in order to enable automated processing. As for ontology descriptions, the W3C recommendation for an ontology language OWL has limitations both on a conceptual level and with respect to some of its formal properties. One proposal for the description of Semantic Web Services is WSML which is the formal language of the WSMO.

The WSML ontology is created on the fly according to the content of the enhanced XML file. The WSML Ontology (see figure 8) mainly consists of: WSML variant, Namespace, A set of non functional properties, Concepts and Axioms.

Concepts and axioms are retrieved stately from XML schema. Table 1 depicts the main transformation rules from XML to WSML language. For example, the XML file introduced above will be translated to the following representation in WSML language.

TABLE: XML TO WSML TRANSLATION RULES

XML Element	WSML Element
Complex_Type Constructor	Concept
Concept Constructor	Concept
Attribute Constructor	Concept
Restriction constructor	axiom
Enumetare constructor	Or operator
Attribute of SimpleConcept	attribute Datatype

Figure 8 presents the translation of the entity Customer of the figure 6. Each attribute is joined to its corresponding value-type by the constructor *OfType*.

```

Concept Customer
Dob OfType date
Address OfType string
  
```

Fig. 8. A screenshot of a the proposed tool

¹⁵ WordNet is an online lexical database designed for use under program control. English nouns, verbs, adjectives, and adverbs are organized into sets of synonyms, each representing a lexicalized concept (See <http://wordnet.princeton.edu>).

However, the non functional properties are introduced by the user by selecting items from a list defined before.

e) Phase 4: Validation. All previous steps may introduce wrong concepts and relationships, thus an automated validation step is needed. This step is often done by hand. Before registering the ontology, it should be validated to ensure WSML correctness and accuracy. If the ontology is not suitable to the WSML language, a domain expert could bring some alterations to correct it. Once the ontology is validated, it will be stored in the .

IV. EXPERIMENTATION

In order to validate and evaluate our approach, a software tool has been created. The execution engine of this tool has been fully implemented in Java with the NetBeans IDE because there is a large amount of libraries that ease the retrieval and parsing of web pages and the construction of ontologies. This tool represents the first step of our efforts in developing a general translation engine for building Semantic Web Service ontology according to the WSMO conceptual model.

The tool is called BUOWES (Building Ontology for Web Service), which is a semi automatic tool for the translation of the WSDL file to an incomplete WSMO ontology.

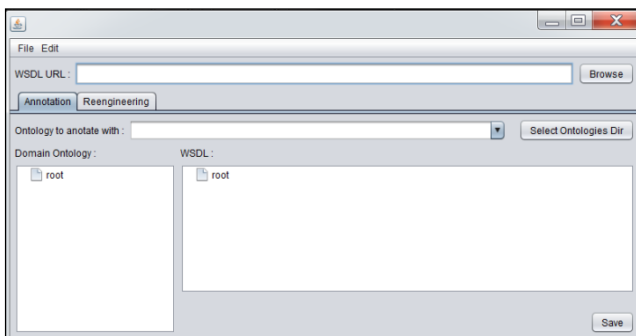


Fig. 9. A screenshot of a the proposed tool

A screen-shot of the BUOWES's user interface is presented in figure 9. BUOWES is a software tool which takes as an input the WSDL specification and domain ontology and returns WSMO ontology description specified in the WSML language as an output.

BUOWES software tool provides a friendly user interface and simple to utilize. You can either upload a WSDL file by its URL or by browsing locally. Then, the wrapper module parses the WSDL file and extracts the XSD (XML Schema Definition) defined between the WSDL type tags.

The mapping engine converts the XSDs extracted to terms used by the ontology, according to a set of mapping rules. The mapping engine produces a list of terms which will be stored in an XML file to be used next.

To improve our proposal with a semantic layer and ensure semantic interoperability, we used a semantic similarity measure, so candidates' concepts of the XML file could be superseded with other ones retrieved by domain ontology. The user participation is fundamental and must be promoted to upload the appropriate domain ontology.

The building module translates the terms into WSML specification and generates the ontology according to the WSML language. The user can visualize the WSMO Ontology as a tree or in an editor specified in the WSML Language.

Finally, the validator module checks the produced ontology to ensure WSML correctness and accuracy.

Non functional properties are added by the user. The user may choose the relevant non functional properties from a list of choices (see figure 10).

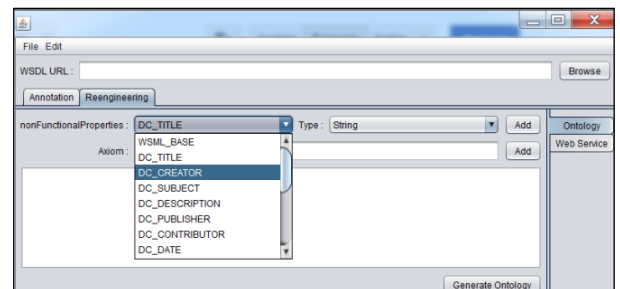


Fig. 10. Adding non-functional properties

We perform our tool on a data set of Web services that contains 20 WSDL File of different domains showing the impact of the proposed tool in decreasing the time and the effort of the building process..

Figure 11 depicts an example of ontology builded for tourism domain.

```

wsmlVariant _"http://www.wsmo.org/wsml/wsml-syntax/wsml-flight"
@namespace {_"http://www.gsno.org/dip/travel/domainOntology#",
dc _"http://purl.org/dc/elements/1.1#",
wsml _"http://www.wsmo.org/wsml/wsml-syntax#"}
@ontology _"http://www.gsno.org/dip/travel/domain.wsml"
+ nfp []
@concept Ticket
+ nfp
dc#description hasValue "concept of a ticket"
endnfp
from ofType Region
to ofType Region
vehicle ofType Vehicle

```

Fig. 11. Building Ontology for a WSDL of Tourism Web Service

Moreover, the experimental results promise that the proposed tool will have a positive impact on the building process as a whole.

V. CONCLUSION

In this paper, we have described the current main Semantic Web Services approaches and we have proposed a novel approach for building Semantic Web

Service ontology according to the WSMO conceptual model.

The proposed technique uses a reverse engineering process and WSDL file of existing Web Service.

Our approach is composed by two main aspects: reverse engineering the WSDL file and engineering the WSMO Ontology specified in the WSML language.

Our approach is based on the idea that semantics of a Web Service can be extracted by analyzing WSDL File. These semantics are supplemented with the domain ontologies and user head knowledge to build the ontology.

Finally, we have implemented a set of end user tools based on web interfaces for:

- The insertion of WSDL Files,
- Calculating the similarity measure,
- Browsing the ontological repository by the final user and producing WSML Ontology.

The strong point of the proposed approach is that it relies on a WSDL File of an existing Web Service without paying attention to the source code of the Web Service application. Also, the use of domain ontology improves semantic interoperability.

The work reported in this paper is a part of large project, and actually, the work still in progress aimed at extending our approach to build Web Services, Goals and Mediators.

We believe that the results of our initial experimentation are quite promising, and we will continue to develop and evaluate this process.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

- [19] H. EL BOUHISSI and M. Malki . Reverse Engineering Existing Web Service Applications", In Proc. Of 16th Working Conference on reverse Engineering,WCRE'09, Published by the IEEE Computer Society, pp 279-283, ISSN Number 1095-1350, ISBN 978-0-7695-3867-9 October 13-16, 2009, Lille, France.
- [20] D. Martin, M. Burstein, J. O. Hobbs, D. Lassila, S. McDermott, S. McIlraith, M. Narayanan, B. Paolucci, T. Parsia, E. Payne, Sirin, N. Srinivasan & K. Sycara (2004). OWL Web Ontology Language for Services (OWL-S), W3C Member Submission.
- [21] J. De Bruijn, C. Bussler, J. Domingue, D. Fensel, M. Hepp, U. Keller, M. Kifer, B. König-Ries, J. Kopecky, R. Lara, H. Lausen, E. Oren, A. Polleres, D. Roman, J. Scicluna, & M.,Stollberg (2005). Web Service Modeling Ontology (WSMO).
- [22] R. E. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, M. Schmidt, A. Sheth & K. Verma, (2005). Web Service Semantics - WSDL-S, W3C Member Submission.
- [23] J. Farrell & H. Lausen (2007). Semantic Annotations for WSDL and XML Schema. W3C Candidate Recommendation, Retrieved from <http://www.w3.org/TR/sawSDL>.
- [24] M. Dimitrov, A. Simov & D. Ognyanov (2005). WSMO Studio – An Integrated Service Environment for WSMO. In Proceedings of the Workshop on WSMO Implementations Innsbruck, Austria, June 6-7.
- [25] M. Kerrigan & A. Mocan (2008). The Web Service Modeling Toolkit. In Proceedings of the 5th European Semantic Web

Conference (ESWC'08), Tenerife, Canary Islands, Spain, June 1-5, pp.812-816.

- [26] W. Xiaomin, A. Murray, Storey, M.-A., & R. Lintern (2004). A Reverse Engineering Approach to Support Software Maintenance: Version Control Knowledge Extraction. In Proceedings of the 11th Working Conference on Reverse Engineering (WCRE'04), Washington, DC, USA, pp. 90-99.
- [27] L. Meng, R. Huang, and J. Gu (2013). A Review of Semantic Similarity Measures in WordNet. International Journal of Hybrid Information Technology. Vol. 6, No. 1.

CS-DSM : Cloud Service Description and Selection Model

Sihem BEY^{*}, Youssef Amghar⁺ and Fahima Nader^{*}

^{*}National School of Computer.ESI. Algeria

⁺INSA-Lyon. LIRIS. France

{s_bey, f_nader}@esi.dz

youssef.amghar@insa-lyon.fr

Abstract—According to statistics, Cloud Computing has experienced a phenomenal propagation in the industry world, in the few last years. Many important enterprises such as Amazon, IBM, Google, etc have adopted it by becoming CSP (Cloud Service Provider). Faced to this huge number of CSPs offering several services with different QoS levels, Cloud Service Discovery and Selection issue becomes more than ever an important researcher's challenge. In this paper, we aim to help consumers to find the best CSP likely to meet his service technical and quality requirements. We propose a Cloud service Description and Selection model CS-DSM that: 1) allows consumer to specify his service functional and non-functional requirements and to evaluate the CSP by providing his feedback; 2) allows CSP to describe the service technical

specifications that he can ensure and to define the QoS that he can guarantee. More precisely, we propose a WSDL (Web Service Description Language) and a UDDI (Universal Description Discovery and Integration) extension models. We propose likewise a Cloud Service Selection Algorithm including a CSPs ranking mechanism based on a qualimetry results.

Index Terms—Cloud Computing, Consumers, Selection Algorithm, Service Description, Service Provider, Qualimetry

ACOHM : Bio-inspired routing protocol for High Mobile Ad Hoc Networks

SAYAD Lamri ^{#1}, AISSANI Djamil [#], BOUALLOUCHE-MEDJKOUNE Louiza [#]

[#]Laboratoire LAMOS, Faculté des Sciences Exactes, Université de Bejaia, ALGERIE

siad_lamri@yahoo.fr

^{*}Département d'Informatique, Université de M'sila, ALGERIE

Abstract— A Mobile Ad Hoc Network (MANET) is set of mobile entities acting both as nodes and routers, and moving without any central administration. The most important advantage of this kind of networks is the short time needed to deploy and thus to use this network. Therefore, MANETs are suitable for rescue and military operations. This paper considers Ad Hoc networks in the context of high mobility. Thus, a new routing protocol, ACOHM, has been proposed to deal with the dynamic aspect of MANETs. The proposed routing protocol takes inspiration from foraging behavior of

ant colonies. In fact, artificial ants are used to mimic this behavior in order to construct multiple routes between a source node and a destination. This approach is, then, coupled with an asynchronous route maintenance mechanism. Simulation results show that our protocol ACOHM outperform AODV in terms of packet delivery ratio and average end-to-end delay.

Keywords — ACO, Bio-inspired, routing, MANET, High Mobility

Routage sémantique d'information dans un réseau p2p de partage de fichier

¹Ahlem Kenniche, ²Lougmiri Zekri, ³Bouziane Beldjilali

Département Informatique, Faculté des Sciences

Université d'Oran Es-Sénia

¹ahlemkenniche@yahoo.fr

²zekri.lougmiri@univ-oran.dz

³bouziane.beldjilali@univ-oran.dz

Abstract- Cet article présente une approche de routage des requêtes dans un réseau pair à pair de partage de fichier en utilisant une méthode fondée sur les profils sémantiques des pairs. Pour cela, chaque pair inclut une matrice de cas et collabore avec les autres pairs pour échanger des requêtes. Une particularité de cette approche est qu'elle permet un enrichissement de la base de cas en utilisant la connaissance et les correspondances relatives aux voisins des pairs. Notre

objectif principal par l'utilisation d'une représentation booléenne de la base de cas est de réduire le temps de calcul lié à une indexation classique et de minimiser l'espace de stockage grâce au stockage binaire. Pour cela, nous explorons plusieurs stratégies complémentaires, notamment pour réduire la taille des index à traiter.

Mots clés : systèmes p2p, routage sémantique, matrice de cas.

The Fusion of SVD and NLS inversions for the Retrieval of a Simulated 3D Urban Structures in a Tomography SAR Application

Hichem Mahgoun

USTHB, Laboratory of Signal and Image Processing, Algiers

Abstract—In this paper we present the model of tomography SAR and the performances of the algorithms used to retrieve thereflectivity signal for a rang azimuth resolution cell. The study is based on two methods, a non parametric estimation method (SVD 1) based on the decomposition in singular value of the projection matrix, and a parametric estimation method (NLS2), based on a non linear square estimator. After a comparison between the performances of the two inversion methods, we apply the chosen methodology to retrieve the reflectivity signal of an elementary structure (cub), and show

the results in radar and ground coordinates. In a second step, we study the behavior of a tomography SAR system for a small town and we apply SVD and NLS inversion separately. Finally, we propose a technique that uses a fusion between SVD and NLS, to achieve rapidity and high accuracy. We conclude about the perspectives of the use of Tomography SAR at large scale for urban area.

Keywords—RADAR, SAR, Tomography, multi-baselines, SVD, NLS, 3D imaging

Les jointures des flux de données dans les réseaux de capteurs sans fil

Djail Boubekour

Ecole Supérieur d'Informatique d'Alger

b_djail@esi.dz

Hidouci Walid-Khaled

Ecole Supérieur d'Informatique d'Alger

wk.hidouci@gmail.com

Abstract - Dans les réseaux de capteurs sans fil, la minimisation des coûts des transmissions demeure un très grand défi. Le défi est d'autant plus grand que la quantité d'information à transmettre est très grande.

Les exécutions des requêtes de jointures constituent un cas où des volumes d'informations très importants doivent être transmis, ce qui entraîne une consommation très élevée de l'énergie fournie par les batteries en place.

Les jointures dans les réseaux de capteurs sont réalisées entre des flux de données. Peu de travaux de recherche ont été

réalisés pour ce type de jointures. La plupart des travaux s'intéressaient aux tables statiques plutôt qu'aux flux de données.

L'article suivant présentera l'état de l'art des techniques suggérées pour l'exécution des requêtes de jointures sur des flux de données.

Keywords: réseaux de capteurs sans fil, coût de communication, jointure intra-réseau, flux de données.

Parallel Pipelined Implementation of DES Cryptographic Algorithm on Multicore Machines

Slimane Mohamed

Industrial Computing and Networking Laboratory
Computer Science Department, University of Oran, BP
1524 Oran, Algeria
Slimanemd@hotmail.com

Sekhri Larbi

Industrial Computing and Networking Laboratory
Computer Science Department, University of Oran, BP
1524 Oran, Algeria
larbi.sekhri@univ-oran.dz

Abstract—The DES (Data Encryption Standard) cipher is considered as the most important symmetric cryptographic algorithm which encrypts the information by blocks of 64 bits length. It is so fast compared to other symmetric algorithms which make it popular as encryption security solution. Its execution on modern multicore platforms does not take great benefit of the power of such machines with its sequential implementation. In order to tackle this problem we must design a parallel version for multicore.

This work tries to make a parallel variant of DES sequential version by processing several blocks in parallel way

in order to speed-up the encryption and decryption process. As the sequential nature of its internal structure, it is more suitable to use pipelined pattern of parallelism to achieve this goal. We consider each step in the algorithm as stage that receives the output of the previous one as its input and process it then send the result to next stage. We will show the benefit of such approach by comparing it with sequential version.

Keywords: DES, symmetric cryptographic algorithm, encryption by bloc, parallel pipelined code, multicore machines.

Election de Leader dans les Réseaux Mobiles Ad Hoc Basée sur le Protocole de Routage TORA

Leila MELIT, Omar YAKHLEF

Département d'informatique
Université de Jijel, Ouled Aissa, Jijel, Algérie

Nadjib BADACHE

Laboratoire des Systèmes Informatiques
USTHB, Alger, Algérie

Résumé— L'élection de leader est un problème fondamental dans les systèmes distribués statiques et dynamiques. La définition classique de ce problème est d'élire finalement un leader unique parmi un ensemble fini de noeuds. Cependant, la résolution du problème d'élection devient plus difficile dans les réseaux mobiles ad hoc où les coupures des liens et la défaillance des noeuds sont considérées comme norme et pas comme exception. Dans cet article, nous proposons un nouvel algorithme d'élection pour les réseaux mobiles ad hoc basé sur le protocole de routage TORA. L'algorithme proposé assure que chaque composante connexe du réseau mobile ad hoc

finira par avoir exactement un leader pour tous ses noeuds. Ainsi, l'algorithme proposé tolère n'importe quel nombre de défaillances de processus et de liens à condition que ce nombre soit fini durant l'exécution de l'algorithme. Nos résultats de simulations montre que notre algorithme permet une réduction de la dissipation d'énergie en minimisant le nombre de messages échangés. En plus, il converge plus vite à un état stable.

Mots clés—élection de leader; réseaux mobiles ad hoc; protocole de routage TORA ; DAG orienté leader ; inversement de liens.

Extraction des relations de causalité dans le domaine médical

Fatma Zohra Belkredim
Département de maths, Université
Hassiba Benbouali
Chlef, Algérie
fzbelkredim@yahoo.fr,

Djamila Hammouche
Département d'informatique,
Université Hassiba Benbouali Chlef,
Algérie
hammouche@yahoo.fr,

Sana Ouraghi
Département d'informatique,
Université Hassiba Benbouali Chlef,
Algérie
souraghi@yahoo.fr

Résumé—L'extraction d'information dans le domaine médical est très importante. Elle permet de fournir des informations et d'en découvrir de nouvelles informations afin de prendre des décisions pour améliorer la qualité des soins. Il s'agit d'extraire automatiquement des informations non structurées sous forme textuelle et de les convertir sous forme structurées en vue d'interrogation de base de données et du traitement automatique du langage naturel.

L'extraction des relations de causalité dans le domaine médical permet d'avoir accès à une information pertinente qui servira aux membres hospitaliers. Ces relations sont exprimées par les concepts qui ont une connotation médicale et causale. Après une introduction, nous présenterons la causalité, son extraction puis son implémentation et la validation.

Mots clés Causalité, Extraction de relations, Domaine médical.

Vers une simulation de l'orchestration des activités mobiles d'apprentissage avec Netlogo

Nassim DENNOUNI, Yvan PETER, Luigi LANCIERI
Équipe NOCE, Laboratoire LIFL
Université Lille 1, France
dennouninas@gmail.com, yvan.peter@univ-lille1.fr /
luigi.lancieri@univ-lille1.fr

Zohra SLAMA
Équipe ISIBA, Laboratoire EEDIS
Université DJILALI LIABES de Sidi Bel Abbes, Algérie
zslama@yahoo.com

Résumé — Pendant le déroulement des sorties pédagogique, les apprenants peuvent créer, partager ou enrichir les activités et les ressources du scénario d'apprentissage. Cependant face à la croissance du nombre de POIs (Point Of Interest), les apprenants sont confrontés au problème de la sélection de l'activité d'apprentissage la plus appropriée par rapport à leurs contextes spatiotemporels. Les systèmes de recommandation basés sur le filtrage collaboratif peuvent être utilisés pour résoudre ce problème. Dans cet article nous allons simuler ce type de réseau d'apprentissage grâce à l'environnement Netlogo. Ce dernier nous a permis de

modéliser notre système de formation qui est composé de plusieurs apprenants agissant en parallèle afin de tester nos différentes techniques d'orchestration des activités mobiles d'apprentissage.

Mots clés — *apprentissage mobile, sortie pédagogique, orchestration des activités mobiles, filtrage collaboratif, Recommandation des POIs, Netlogo.*

ICNTC'2015

Chlef, Algérie
3-4 Mars, 2015

ICNTC'2015 a relevé les défis liés à la construction d'une future plateforme pour internet et à l'utilisation des nouvelles technologies de la communication qui sont basées sur des infrastructures fixes, sans fil et mobiles. Cette manifestation a été aussi un lieu idéal pour les chercheurs, les institutions officielles, les décideurs, les participants du secteur public et les industriels car elle a facilité le transfert et le partage des résultats de recherches au sein de la communauté des chercheurs et des doctorants.

