

5^{ème} EDITION

CNRIA'2013

Du 24 au 27 avril 2013



Ziguinchor
(SENEGAL)

PROCEEDINGS(ACTES)

COLLOQUE NATIONAL SUR LA RECHERCHE EN
INFORMATIQUE ET SES APPLICATIONS

organisé par l'ASCII (Association Sénégalaise des Chercheurs en Informatique)



Préface

Le Colloque National sur la Recherche en Informatique et ses Applications (CNRIA) a connu, depuis sa création, quatre (04) éditions qui se sont tenues respectivement en 2005, 2010, 2011 et 2012. Ces différentes éditions ont compté en moyenne une vingtaine de communications dans le domaine des réseaux, des bases de données, des systèmes distribués, de la modélisation, du web sémantique, de la fouille de données, etc. Le CNRIA est devenu aujourd'hui un évènement scientifique important pour les chercheurs informaticiens sénégalais et leurs partenaires sous-régionaux et internationaux. Il favorise la synergie entre les chercheurs et leur offre un moyen de discuter de leurs travaux avec leurs pairs et, par conséquent, éviter l'isolement. Il faut dire que les chercheurs en informatique en activité au Sénégal sont encore relativement peu nombreux, jeunes pour la plupart et dispersés au sein des cinq (5) universités que comptent le pays.

Pour la présente édition, seize (16) articles ont été retenus par le comité scientifique : huit (8) dans le domaine des réseaux et systèmes distribués, trois (3) dans le domaine des ontologies et du web sémantique, trois (3) dans le domaine de l'ingénierie logicielle, des agents et des services et deux (2) dans le domaine des bases de données et du datamining. Tous les articles reçus ont fait l'objet de deux, trois, voire quatre relectures. Nous avons aussi 2 posters et 2 démonstrations.

Deux conférences invitées portant sur les réseaux de capteurs sont aussi retenues. Elles seront données par Cong Duc PHAM, Professeur à l'Université de Pau et des Pays de l'Adour et Bernard Pottier, Professeur à l'Université de Bretagne Occidentale.

Il y'a plus de soixante-dix (70) participants inscrits en provenance du Sénégal, de la Mauritanie, de la Côte d'Ivoire, de la France et de la Suisse. Il s'agit, principalement, d'enseignants-chercheurs et doctorants des universités sénégalaises, de doctorants sénégalais préparant leur thèse en France et de chercheurs partenaires invités.

Cette édition est celle du lancement de l'Association Sénégalaise des Chercheurs en Informatique (ASCII), un « bébé » du CNRIA qui aura, entre autres missions, celle d'assurer la pérennité du Colloque. Aussi, des ateliers sur l'harmonisation des programmes d'informatique au sein des universités sénégalaises,

l'orientation et l'organisation de la recherche en informatique se tiennent en marge du Colloque.

Nous remercions tous les chercheurs qui ont répondu à notre appel à candidatures avec une mention spéciale pour les conférenciers invités.

Nous félicitons tous les auteurs qui ont vu leur(s) article(s) retenu(s) par le comité scientifique et remercions par la même occasion tous les membres du comité scientifique qui n'ont ménagé aucun effort pour effectuer une relecture des articles qui leur ont été affectés.

Nous terminerons par remercier tous ceux qui ont permis, de près ou de loin, la tenue de cette édition qui se déroule dans la belle région de Casamance. Nous accordons une mention spéciale au Ministre de l'Enseignement supérieur et de la recherche, au Directeur de l'Institut français du Sénégal, aux Recteurs des Universités de Ziguinchor, de Dakar, de Saint-Louis, de Bambey et de Thiès. Nous associons à ces remerciements tous les responsables de structure (départements, UFR, facultés, instituts) qui ont bien voulu soutenir l'organisation de CNRIA'2013. Pour finir, nous tenons à féliciter très chaleureusement le comité d'organisation local, dirigé par Dr Youssou DIENG qui, malgré toutes les difficultés rencontrées, a réussi à organiser cette édition qui, sans nul doute, marquera l'esprit des membres de la communauté CNRIA. Bonne lecture !

Pour Le comité de pilotage de CNRIA'2013

Dr Youssou DIENG (UAS)

Pr. Ibrahima NIANG (UCAD)

Pr. Moussa LO (UGB)

Comité de pilotage national

Ousmane SALL	Université de Thiès
Dame SAMB	Université de Thiès
Cheikh BA	Université Gaston Berger de Saint-Louis
Gaoussou CAMARA	Université de Bambey
Adrien BASSE	Université de Bambey
Youssou KASSE	Université de Bambey
Mouhamadou THIAM	Université de Thiès
Maïssa MBAYE	Université Gaston Berger de Saint-Louis
Jean Marie DEMBELE	Université Gaston Berger de Saint-Louis
Moussa LO	Université Gaston Berger de Saint-Louis
Samba NDIAYE	Université Cheikh Anta Diop de Dakar
Ibrahima NIANG	Université Cheikh Anta Diop de Dakar
Ousmane THIARE	Université Gaston Berger de Saint-Louis
Bamba GUEYE	Université Cheikh Anta Diop de Dakar
Idrissa SARR	Université Cheikh Anta Diop de Dakar
Gervais MENDY	Université Cheikh Anta Diop de Dakar
Khadiatou WANE KEITA	Université Cheikh Anta Diop de Dakar
Amadou GAYE	Université Cheikh Anta Diop de Dakar

Comité scientifique

Arnaud GIACOMETTI	Université de Tours
Alassane DIOP	Télé-Université du Québec
Almamy KONTE	Université Cheikh Anta Diop de Dakar
Amadou COULIBALY	INSA Strasbourg
Abdourahmane FAYE	SAP Suisse
Amrane HOCINE	Université de Pau et des Pays de l'Adour
Bamba GUEYE	Université Cheikh Anta Diop de Dakar
Béatrice BOUCHOU	Université de Tours
Borilli M. Jonas SOME	Ecole Polytechnique de Bobodioulasso
Catherine Faron ZUCKER	Université de Nice
Cheikh BA	Université Gaston Berger de Saint-Louis
Cheikh SARR	Université de Thiès
Cheikh Talibouya DIOP	Université Gaston Berger de Saint-Louis
Chérif DIALLO	Université Gaston Berger de Saint-Louis
Demba COULIBALY	Université de Bamako
Dominique LAURENT	Université de Cergy-Pontoise
Fatou Kamara SANGARE	Université Gaston Berger de Saint-Louis
Gayo DIALLO	Université Bordeaux Segalen
Ibrahima NGOM	Université Cheikh Anta Diop de Dakar
Ibrahima NIANG	Université Cheikh Anta Diop de Dakar
Idrissa SARR	Université Cheikh Anta Diop de Dakar
Jean Marie DEMBELE	Université Gaston Berger de Saint-Louis
Kader KEITA	Université de Bamako
Khadidiatou Wane KEITA	Université Cheikh Anta Diop de Dakar
Khalifa GAYE	Université de Ziguinchor
Maïssa MBAYE	Université Gaston Berger de Saint-Louis
Mamadou BOUSSO	Université de Thiès
Mamadou Youry SALL	Université Gaston Berger de Saint-Louis
Marcellin Konan BROU	Institut National Polytechnique de Yamoussoukro
Marie NDIAYE	Université de Ziguinchor
Maurice TCHUENTE	Université de Yaoundé I
Mouhamadou THIAM	Université de Thiès
Moussa LO	Université Gaston Berger de Saint-Louis
Nacera BENACER	Supelec, Paris

Nathalie PERNELLE	LRI Paris
Ndeye Fatou NGOM	Université Cheikh Anta Diop de Dakar
Olivier CORBY	INRIA Sophia Antipolis
Olivier FLAUZAC	Université de Reims Champagne Ardenne
Olivier MONGA	IRD Yaoundé
Oumar DIOP	Université Gaston Berger de Saint-Louis
Ousmane SALL	Université de Thiès
Ousmane THIARE	Université Gaston Berger de Saint-Louis
Patrick MARCEL	Université de Tours
Rédouane LBATH	Université de Toulouse 1 Capitole
Rim DJEDIDI	Lim&Bio Paris 13
Sadouaouan MALO	Ecole Polytechnique de Bobodioulasso
Samba DIAW	Université de Ziguinchor
Samba NDIAYE	Université Cheikh Anta Diop de Dakar
Seydina Moussa NDIAYE	Université Gaston Berger de Saint-Louis
Sylvie DEPRES	Lim&Bio Paris 13
Thierno TOUNKARA	Telecom Bretagne
Yahya SLIMANI	Université de Tunis

Table des matières

Résumés des conférenciers invités	2
Aspects spatiaux temporels et simulation des réseaux de capteurs <i>Bernard POTTIER, Professeur à l'université de Bretagne Occi-</i> <i>dentale</i>	3
Objets communicants et réseaux de capteurs pour une intelligence dis- tribuée au service de la société <i>Cong Duc PHAM, Professeur à l'université de Pau et des Pays</i> <i>de l'Adour</i>	4
Articles Longs	5
Régulation du débit des flux best effort dans les réseaux ad hoc mul- tisauts <i>Cheikh SARR</i>	6
SENTINEL : un mécanisme pour une prolongation de la durée de vie des réseaux de capteurs sans fil <i>Dame DIONGUE et Ousmane THIARE</i>	14
Softcust : A Friendly Localized Platform for Multilingual Semantic Communication <i>Fabian CRETTON, Zhan LIU and Anne LE CALVE</i>	22
Vers un système de veille épidémiologique à base d'ontologie <i>Gaoussou CAMARA, Sylvie DESPRES, Rim DJEDIDI, Moussa</i> <i>LO</i>	30
Gestion des services dans P2P4GS <i>Bassirou GUEYE, Olivier FLAUZAC, Cyril RABAT et Ibrahima</i> <i>NIANG</i>	38
Minimisation de la Puissance Totale dans les Systèmes Multi-Utilisateurs OFDM avec l'Algorithme Génétique <i>Abdourahmane NDIAYE, Samuel OUYA, Gervais MENDY et S.</i> <i>Mohamed FARSSI</i>	46
Réutilisation de ressources terminologiques existantes pour la construc- tion d'ontologie de domaine <i>Khadim DRAME, Gayo DIALLO, Fleur MOUGIN</i>	54
Amélioration de la consommation d'énergie fondée sur la redondance de mesure dans les WSNs <i>Ibrahima DIANE, Rahim KACIMI, Zoubir MAMMERI et Ibra-</i> <i>hima NIANG</i>	62
Factorisation multi-biais pour de meilleures recommandations <i>Modou GUEYE, Talel ABDESSALEM et Hubert NAACKE</i>	70

Exploration des graphes dynamiques T-intervalle-connexes <i>Ahmed Mouhamadou WADE</i>	78
Étude comparative entre solutions de clustering auto-stabilisantes à k sauts dans les réseaux Ad Hoc <i>Mandicou BA, Olivier FLAUZAC, Rafik MAKHLOUFI, Florent NOLOT et Ibrahima NIANG</i>	86
Amélioration de l'Algorithme P.E.G par Minimisation des Trapping-Set TS(5,3) sur les Codes LDPC <i>Madiagne DIOUF, David DECLERCQ et Samuel OUYA</i>	94
Handling of Shared Model Elements in MDE Software Processes <i>Ibrahima FALL, Reda BENDRAOU, Marie-Pierre GERVAIS et Xavier BLANC</i>	102
Evaluation and comparison of RSSI based indoor localization algorithms <i>Ndeye Amy DIENG, Claude CHAUDET, Laurent TOUTAIN et Tayeb BEN MERIEM</i>	110
Architecture d'un système multi-agents sémantique : Application au domaine changement climatique et vulnérabilité urbaine <i>Ibrahima DIOP, Moussa LO, Jean Marie DEMBELE et Papa Alioune CISSE</i>	118
Traitement décentralisé de requêtes de biodiversité <i>Ndiouma BAME, Hubert NAACKKE, Idrissa SARR et Samba NDIAYE</i>	126
Posters	137
3D image reconstruction using Java 3D <i>Ndèye Fatou NGOM, Chérif NDIAYE et Oumar NIANG</i>	138
Modélisation à base d'agents de la propagation de la bilharziose <i>Papa Alioune CISSE, Jean Marie DEMBELLE, Moussa LO et Christophe CAMBIER</i>	139
Démonstrations/Tutoriels	141
Bibliothèques Java de manipulation d'images en VITA (Vision artificielle et traitement d'images) <i>Ndèye Fatou NGOM</i>	142
Introduction à la programmation des capteurs sans fil <i>Mahamadou TRAORE, Bernard POTTIER et Ousmane THIARE</i>	144
Index des auteurs	147

Résumés des conférenciers invités

Aspects spatiaux temporels et simulation des réseaux de capteurs

Bernard POTTIER, Professeur à l'université de Bretagne Occidentale

Résumé. Les réseaux de capteurs sans fil (RdC) permettent de prélever des informations physiquement distribuées, de calculer des diagnostics, et de prendre des décisions, quelquefois par des moyens centralisés, quelquefois localement, sur les sites d'observation. Les instants des prises de mesure, le temps de calcul du diagnostic, et de mise en œuvre d'une réaction peuvent avoir une importance primordiale pour une application critique. Pour des systèmes dédiés, une construction verticale de systèmes de capteurs *partageant une référence temporelle* apparaît viable, sinon souhaitable. Cette construction peut aussi se justifier par la nécessité de coupler l'outil d'observation RdC et le processus physique ou social contrôlé afin d'en valider les fonctions (systèmes cyber-physiques). Nous décrivons une chaîne d'outils et ses transitions allant de la saisie géo-localisée *à-la-google map* à la synthèse de code pour les microcontrôleurs des capteurs. La partie haute du flût permet la saisie et la localisation des capteurs dans une situation géographique. L'analyse de cette distribution permet une synthèse spéculative d'un graphe de processus sur lequel on applique des comportements distribués. Nous montrons deux techniques de simulation parallèles, synchrones, sur GPU Nvidia, et par réseaux de processus communiquant, et nous discutons les couplages avec les interfaces de présentation graphique, par accès mémoire pour le GPU, et par multiplexeur pour les processus, avec une application au réseau *son* de Smart Santander.

Objets communicants et réseaux de capteurs pour une intelligence distribuée au service de la société

Cong Duc PHAM, Professeur à l'université de Pau et des Pays de l'Adour

Résumé. Un capteur est un petit appareil autonome capable d'effectuer des mesures simples sur son environnement immédiat. L'utilisation de ces capteurs n'a rien d'une nouveauté, ceux-ci sont utilisés depuis longtemps dans des domaines comme l'aéronautique ou l'automobile. Ce qui est novateur, c'est la possibilité pour ces capteurs de communiquer par ondes radio (réseaux sans fil) avec d'autres capteurs proches (quelques mètres) et pour certains d'embarquer de la capacité de traitement (processeur) et de la mémoire. On peut ainsi constituer un réseau de capteurs sans fil (RCSF) qui collaborent sur une étendue assez vaste. Sous la dénomination de l'Internet de Objets, ces possibilités de communications et de traitement peuvent également être intégrées de manière plus vaste dans un grand nombre d'objets et d'appareils de la vie courant. Même si aujourd'hui les domaines des réseaux de capteurs et celui de l'Internet de Objets restent assez distinct l'un de l'autre, à terme, il est sans doute prévisible que ces 2 technologies vont être utilisées de manière complémentaire pour bâtir une intelligence distribuée et pervasive pour un grand nombre d'applications civiles. Dans cet exposé, nous présenterons tout d'abord les 2 technologies réseaux de capteurs et Internet des Objets en indiquant les spécificités propres à chacune ; nous présenterons quelques projets concrets mettant en scène ces technologies pour des applications civiles : réseaux de surveillance, logistique, villes intelligentes,... ; et surtout, nous essayerons de présenter les verrous et les défis scientifiques qui font que ces 2 domaines de recherche sont particulièrement fédérateurs pour rassembler un grand nombre de compétences informatiques.

Articles Longs

1. Introduction

Les réseaux ad hoc sont des réseaux autonomes ne possédant aucune infrastructure fixe. L'absence d'une gestion centralisée des communications dans ce contexte ad hoc rend difficile le partage des ressources entre les différents flux. Etant données les performances de ces réseaux, il semble utile de différencier les applications : celles qui exigent des garanties en terme de bande passante, de délai ou d'une métrique quelconque, communément appelés flux *QoS*, et d'autres qui sont beaucoup plus tolérantes à la dégradation de leurs ressources appelées applications *Best Effort*.

Beaucoup de travaux se sont essentiellement concentrés sur la garantie de ressources pour les applications QoS sans se soucier de la présence des flux Best Effort [1, 2, 3]. Cette absence de considération peut rapidement conduire à des configurations où la majeure partie des ressources disponibles est occupée par ces flux Best Effort.

Dans cet article, nous proposons le protocole DRBT pour *Dynamic Regulation of Best Effort Traffic*. L'idée principale de cette approche consiste à estimer de manière différenciée la bande passante résiduelle des liens afin de diminuer si nécessaire le débit des flux Best Effort. Cette diminution permet d'augmenter le taux d'acceptation des flux QoS existants où à venir. L'utilisation du canal radio devant être optimale, lorsque les flux QoS s'arrêtent de transmettre ou se déplacent, les flux Best Effort doivent alors dynamiquement augmenter leur débit jusqu'à leur valeur initiale. La technologie sous-jacent pour les couches physique et MAC est le protocole IEEE 802.11 [7].

2. Etat de l'art

Dans [6], une solution proposée consiste à séparer les trafics en utilisant deux canaux de transmissions distincts. La difficulté de conception de cette technique est un obstacle majeur et rend son utilisation peu probable surtout dans un contexte ad hoc.

SWAN [5] est un protocole distribué qui n'utilise aucun messages de contrôle pour garantir dynamiquement la bande passante des flux QoS. Pour atteindre cet objectif SWAN met en place trois mécanismes. Un mécanisme de contrôle d'admission à la source des flux QoS et deux mécanismes de régulation dynamique de trafic, l'un pour les flux QoS et l'autre pour les flux Best Effort. Le contrôle d'admission à la source utilise une approche intrusive pour estimer la bande passante résiduelle. Avant chaque transmission d'un flux QoS, une sonde est envoyée de la source vers la destination pour évaluer la bande passante résiduelle le long de ce chemin. En fonction de cette valeur, le contrôle d'admission à la source décide de l'envoi ou non du flux QoS concerné. Il n'y a pas de routage car le chemin entre la source et la destination est supposé connu.

Lorsque le protocole SWAN estime que le réseau est dans un état de congestion, alors les mécanismes de régulations sont déclenchés.

Le protocole QPART [4] se base sur une estimation passive de la bande passante et un mécanisme de régulation dynamique du débit des flux Best Effort. Il classe les trafics avec qualité de service en fonction de deux métriques : le délai et la bande passante.

Toutefois, nous nous intéresserons ici qu'à la bande passante résiduelle comme métrique, puisque c'est celle que nous avons considérée dans cette section.

L'estimation de la bande passante restante effectuée dans QPART se base sur la théorie des files d'attente et plus précisément sur le protocole RED. L'algorithme RED démontre que pour maintenir un débit *constant* sur un chemin multi-saut, la taille des files d'attentes doit être maintenue en dessous d'un certain seuil prédéterminé. Si ce seuil est dépassé alors les mécanismes de régulation sont automatiquement démarrés afin de résorber le surplus de trafic.

En cas de congestion, la taille de la fenêtre de contention des flux QoS et Best Effort est mise à jour de manière à rendre la transmission des flux QoS prioritaire. Parallèlement, l'algorithme de QPART sélectionne des flux susceptibles d'être rejetés en se basant sur une priorité des flux.

Cependant, QPART souffre des mêmes limitations que le protocole SWAN. Premièrement, il se base sur un mécanisme d'estimation de la bande passante utilisé dans les réseaux filaires et qui est inadapté dans un contexte ad hoc. En effet l'estimation utilisée dans QPART se base sur la théorie de file d'attente. L'algorithme RED utilisé stipule que pour maintenir un débit constant sur un chemin multi-saut, on doit maintenir la taille des files d'attentes au niveau de chaque émetteur en dessous d'un seuil déterminé. Cependant, dans un contexte ad hoc sans fil, cette assertion n'est pas toujours vérifiée. En effet, il existe de nombreuses topologies, notamment celle des stations cachées où les files d'attente des restent en dessous d'un seuil mais les collisions à la réception engendrent des retransmissions fréquentes.

Deuxièmement, le mécanisme de régulation dynamique du trafic Best Effort dans le protocole QPART, consistant à faire varier la taille de la fenêtre de contention, ne permet d'augmenter ou de diminuer le débit des ces flux Best Effort qu'en moyenne. Une variation de la taille de la fenêtre de contention ne permet pas de prédire avec précision quelle sera la variation la bande passante associée. Cela s'explique simplement par le fait qu'au niveau du protocole IEEE 802.11, la transmission sur le canal radio dépend aussi bien de la taille de la fenêtre de contention mais également de l'état du canal (occupé ou libre) et des collisions qui surviennent à la réception. C'est pourquoi QPART ne peut garantir avec précision la bande passante des flux QoS ni garantir une utilisation au mieux du médium pour les flux Best Effort.

3. Gestion du débit des flux Best Effort

Dans cette section, nous présentons en détail l'estimation de la bande passante résiduelle des flux QoS et comment cette estimation est utilisée pour réguler le débit des flux Best Effort.

3.1. Estimation de la bande passante résiduelle pour les flux QoS

La première étape, afin de garantir le débit des flux QoS, est d'estimer la bande passante résiduelle que l'on peut leur allouer. L'estimation doit être distribuée et prendre en compte les différents phénomènes propres aux réseaux ad hoc (interférences, collisions etc.). Le protocole ABE [1] fournit une estimation fiable de la bande passante résiduelle

SARR

des liens radio. Le principe de base de ce protocole est que les monitorent le support radio afin de connaître le taux d'occupation du canal radio. Par la suite, une estimation de la synchronisation des mobiles émetteur et récepteur, ainsi que du taux de collision sont effectuées afin de déterminer la bande passante résiduelle des liens. Pour plus de détails consulter [1]. Toutefois, l'estimation dans ABE ne prend pas en compte le type (QoS ou Best Effort) des transmissions dans le réseau. Nous rajoutons donc au protocole ABE, un mécanisme afin de différencier les paquets.

Cette différenciation s'effectue au niveau de la couche MAC et consiste à ne mesurer que l'occupation médium des paquets QoS durant la phase d'écoute du support radio. En effet, lorsqu'un paquet est transmis sur le médium radio, les qui écoutent le support à portée de communication peuvent identifier la nature du paquet (QoS ou Best Effort) en analysant les champs ToS de l'en tête IP. Un champ ToS = 1 correspond à un paquet QoS.

Cependant, dans un réseau ad hoc, les paquets se trouvant dans la zone de détection de porteuse ne sont pas décodables. Dans ce cas précis, cette différenciation ne peut se faire. Toutefois, pour se placer dans le pire des cas, nous considérons tous les paquets se trouvant dans la zone de détection de porteuse comme des paquets QoS. En d'autres termes, les transmissions Best Effort se trouvant dans la zone de détection de porteuse sont considérés comme des flux QoS.

Dans la suite de cet article, le terme *bande passante résiduelle différenciée* correspond à la bande passante résiduelle estimée en ne prenant en compte que les transmissions QoS.

3.2. Régulation du débit des flux Best Effort

L'évaluation précédente ne signifie pas que le flux QoS peut émettre au débit calculé puisque la bande passante libre évaluée ne prend pas en compte les éventuels flux Best Effort existants. Par conséquent, il est nécessaire de réguler le débit des flux Best Effort en fonction de l'estimation de la bande passante résiduelle des flux QoS présentée dans la section précédente. Dans DRBT, cette régulation se déroule en deux étapes :

1) Réduire, si nécessaire, le débit des flux Best Effort lorsqu'un nouveau flux QoS désire transmettre et ne trouve pas assez de bande passante résiduelle, car une partie de celle-ci est occupée par ces trafics Best Effort.

2) Augmenter le débit de ces flux Best Effort lorsqu'un flux QoS libère de la bande passante ou se déplace dans une autre zone de transmission

3.2.1. Réduction du débit des flux Best Effort

Le protocole DRBT n'utilise que les paquets de contrôle classiques tels que les RREQ (Route Request) et RREP (Route Reply) que l'on retrouve dans la plupart des protocoles de routage réactifs. Ainsi, pour chaque nouvelle transmission d'un flux QoS, un paquet de RREQ est envoyé au préalable afin de réserver les ressources ainsi que le paquet de réponse RREP correspondant. Les informations stockées dans ces paquets sont :

- Le débit désiré par le nouveau flux QoS (*DebitQoS*).
- Le nombre de flux Best Effort se trouvant dans le voisinage du flux QoS concerné (*nbBE*). En effet, chaque flux Best Effort possède un identifiant unique propagé à travers les paquets *Hello*. En conséquence, chaque peut estimer le nombre de flux Best Effort dans son voisinage en analysant ses identifiants.

La précision de notre protocole DRBT passe aussi bien par une estimation précise de la bande passante résiduelle différenciée que par un processus de routage efficace. Nous utilisons un routage réactif proche de celui d'AODV. La source d'un flux QoS envoie un paquet de RREQ à l'ensemble de ses voisins. Ce paquet contient en plus des deux champs décrits précédemment, l'adresse de l'émetteur, la destination et un numéro de séquence unique. Tout recevant ce paquet de RREQ effectue un simple contrôle d'admission en vérifiant si la bande passante désirée, indiquée au niveau du paquet de RREQ, est inférieure à la valeur de la bande passante résiduelle différenciée sur ce lien. Si c'est le cas, le ajoute son adresse dans la route, met à jour la valeur de $nbBE$ en ajoutant le nombre de flux voisins Best Effort (différents de ceux du voisin précédent ayant envoyé le RREQ) et retransmet ce paquet de RREQ. Ce processus continue jusqu'au destinataire qui après avoir effectué le dernier contrôle d'admission, envoie le paquet de RREP en unicast, par le chemin inverse afin de réserver les ressources.

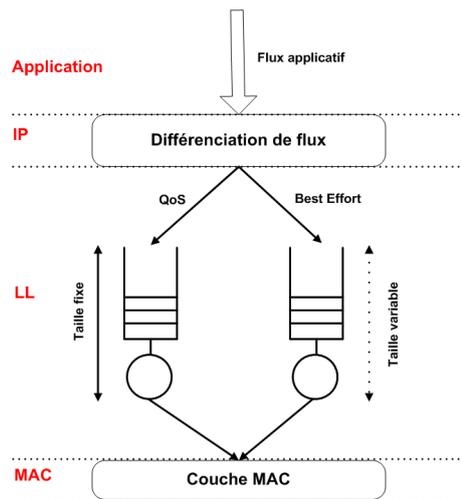


Figure 1. Architecture interne d'un nœud DRBT

A chaque réception d'un paquet de RREQ ou de RREP par un émetteur d'un flux Best Effort, ce dernier estime s'il y a assez de bande passante résiduelle pour transporter ce flux QoS sans dégrader son débit. Si ce n'est pas le cas, il réduit son débit en envoyant un paquet appelé DRP pour *Dynamic Regulation Packet*. Ce paquet, envoyé depuis la couche IP vers la couche LL du même nœud, ne circule pas à travers le réseau. Les informations stockées dans ce paquet DRP sont : la valeur de la bande passante désirée par le flux QoS (*DebitQoS*) et extraite du paquet de RREQ ou RREP et la valeur actuelle de débit du flux Best Effort (*DebitBE*). Enfin, lorsque la couche LL réceptionne ce paquet DRP, elle enclenche le mécanisme de réduction du débit des flux Best Effort.

En pratique, deux files d'attente virtuelles sont utilisées. La première sert à transporter les paquets QoS et la seconde les paquets Best Effort comme indiqué sur la figure(1). Pour diminuer de manière précise le débit du trafic Best Effort, nous réduisons la taille de

SARR

la file Best Effort jusqu'à une valeur seuil. Cette valeur seuil, notée *Seuil*, varie à chaque réception d'un nouveau paquet DRP. Elle est obtenue grâce aux formules :

$$Seuil = \frac{DebitBE}{BPbesteffort} \quad [1] \quad BPbesteffort = \frac{BPDiff - DebitQoS}{nbBE} \quad [2]$$

D'après l'équation 1, si la valeur du seuil est supérieure à un, alors le débit du trafic Best Effort doit être diminué. L'équation 2 représente alors la nouvelle bande passante maximale à allouer au flux Best Effort, si le nouveau flux QoS est accepté.

Une fois que le seuil est déterminé, la nouvelle taille de la file d'attente des flux Best Effort est simplement obtenue en faisant le ratio entre le nombre de paquets Best Effort entrés dans la file d'attente durant une période de mesure d'une seconde et le Seuil dans l'équation 1. Nous avons choisi de fixer cette période à une seconde car au niveau du protocole ABE, les débits de l'ensemble des liens du réseau sont mis à jour toutes les secondes.

3.2.2. Augmentation du débit des flux Best Effort

Lorsqu'un flux QoS arrête sa transmission ou se déplace dans une autre zone de communication, tous les flux Best Effort ayant réduit leur débit à cause de la présence de ce flux QoS, peuvent augmenter de nouveau leur débit jusqu'à la valeur initiale. Cette augmentation permet d'utiliser de manière optimale et dynamique les ressources disponibles qui se libèrent dans le réseau. Pour atteindre cet objectif nous utilisons encore une fois les paquets *Hello*. Chaque transportant un flux QoS stocke dans ces paquets *Hello*, des informations sur l'identité de ce flux et la valeur de la bande passante résiduelle différenciée.

Quand un flux QoS arrête de transmettre ou libère de la bande passante, ce changement d'état est indiqué au niveau des paquets *Hello*. L'émetteur d'un flux Best Effort qui se trouve dans le voisinage de ce flux QoS est en mesure d'intercepter ces paquets *Hello* indiquant qu'un flux QoS a libéré de la bande passante. Finalement, le flux Best Effort peut augmenter son débit en fonction de la nouvelle valeur de la bande passante disponible. En cas de mobilité de certains du flux, les voisins émetteurs de flux Best Effort ne reçoivent alors plus de paquets *Hello* de ces derniers. En conséquence, ils augmentent leur débit jusqu'à la dernière valeur qui avait été conservée.

4. Evaluation des performances

A travers le simulateur NS-2, nous avons évalué les performances de DRBT. Nous comparons les résultats obtenus avec trois autres protocoles : le protocole au mieux AODV [8] qui fournit un routage basique sans contrôle d'admission ; le protocole ABE [1] dont la comparaison va nous permettre de voir quel est l'apport d'une différenciation de flux lorsqu'on régule le débit des communications Best Effort ; le protocole QPART [4] que nous avons implémenté et qui servira de base pour la comparaison avec DRBT lorsqu'une différenciation est réalisée.

Nous introduisons une nouvelle métrique (ϕ) représentant le taux d'acceptation des flux QoS et défini par :

$$\phi = \frac{\text{Nombre de flux QoS admis correctement}}{\text{Nombre total de flux QoS dans le réseau}} \quad [3]$$

Un flux QoS admis correctement est un flux QoS n'ayant pas subi plus de 5% de dégradation de son débit lors de sa transmission. Cette situation implique donc que l'estimation de la bande passante résiduelle différenciée et la phase de contrôle d'admission soient fiables. Cette métrique va également nous permettre d'estimer la fiabilité de l'estimation différenciée de la bande passante résiduelle. En effet, une estimation erronée entraînerait irrémédiablement une dégradation du débit des flux QoS et par conséquent une diminution de la valeur de la variable ϕ .

Nous avons procédé à des simulations pour estimer la valeur du paramètre ϕ . Le nombre de dans le réseau considéré varie de 10 à 50 et sont positionnés aléatoirement sur un carré de 1000m x 1000m. La portée de communication est de 250m tandis que la portée de détection de porteuse est de 550m. Dix flux CBR, dont cinq Best Effort et cinq QoS sont établis entre des source et destination choisis aléatoirement. Les débits sont distribués uniformément dans l'intervalle [0-100] Kb/s. Chaque simulation dure 100 secondes et les résultats présentés constituent la moyenne de 30 simulations pour un nombre de défini.

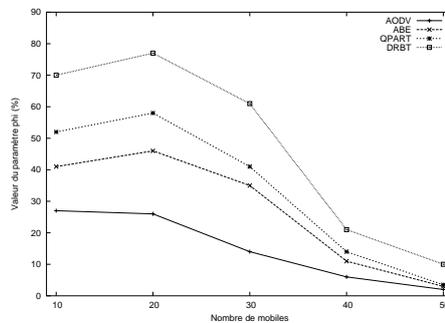


Figure 2. Taux d'acceptation des flux QoS avec les protocoles AODV, ABE, QPART et DRBT

La figure 2 représente la valeur du paramètre ϕ en fonction du protocole utilisé. De manière évidente, plus le réseau devient dense, plus le taux d'acceptation ϕ des flux QoS diminue car la bande passante résiduelle des liens devient plus faible. Lorsque le réseau est peu dense (entre 10 et 20 nœuds), le taux d'acceptation ϕ est relativement élevé pour notre protocole (70%) tandis que le protocole QPART achemine environ 51% des flux QoS. Ainsi, les deux mécanismes de différenciation de flux et d'estimation différenciée de la bande passante résiduelle permettent d'augmenter le taux d'acceptation des flux QoS. Cependant, lorsque le réseau est moyennement dense (entre 20 et 30 nœuds), le taux d'acceptation des flux QoS de tous les protocoles commencent à diminuer. Toutefois le protocole DRBT arrive encore à acheminer jusqu'à 61% des flux QoS présents. Enfin, lorsque le réseau est très dense (entre 40 et 50 nœuds), la bande passante résiduelle des

SARR

liens devenant très faible, une diminution du débit des flux Best Effort est insuffisante pour garantir les ressources au flux QoS. Cependant avec DRBT, 10% des flux QoS sont encore acheminés avec les conditions requises alors que les autres protocoles acheminent dans le meilleur des cas 3.5% de ces flux.

Références

- [1] CHEIKH SARR, CLAUDE CHAUDET, GUILLAUME CHELIUS, ISABELLE GUÉRIN-LASSOUS, « Bandwidth Estimation for IEEE 802.11-based Ad Hoc networks », *IEEE International Journal of Transaction on Mobile Computing (TMC)*, vol. 7, n° 10, Octobre 2010
- [2] HAITAO ZHAO, SHAN WANG, JIBO WEI, AN SONG, YING LI, « Model-based approach for available bandwidth prediction in multi-hop wireless networks », *SCIENCE CHINA Information Sciences*, vol. 54, n° 9, pp. 1916-1927, Octobre 2011
- [3] CLAUDE CHAUDET, ISABELLE GUÉRIN LASSOUS, « BRuIT - Bandwidth Reservation under InTerferences influence », *Proceedings of European Wireless 2002 (EW2002)*, n° Oct 2002
- [4] YALING YANG AND ROBIN KRAVETS, « Distributed QoS Guarantees for Realtime Traffic in Ad Hoc Networks », *IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, n° pp. 118-127, Oct 2004
- [5] GAHNG-SEOP AHN, ANDREW T. CAMPBELL, ANDREAS VERES AND LI-HSIANGSUN, « SWAN : Service Differentiation in Wireless Ad Hoc Networks », *IEEE INFOCOM*, n° 2002
- [6] GAHNG-SEOP AHN, ANDREW T. CAMPBELL, ANDREAS VERES AND LI-HSIANGSUN, « Towards a fully distributed QoS-aware MAC protocol for multihop wireless networks », *International Workshop on Wireless Ad-hoc Networks*, n° 2005
- [7] MÜHLETHALER PAUL, « 802.11 et les réseaux sans fil », *Chapitre de livre - Paris : Eyrolles*, n° 2002
- [8] CHARLES E. PERKINS, ELIZABETH M. BELDING-ROYER AND SAMIR DAS, « Ad hoc On-Demand Distance Vector (AODV) Routing. Internet Request For Comments RFC 3561 », *Internet Engineering Task Force IETF*, n° July 2003

1. Introduction

Les récentes avancées technologiques dans le domaine des communications sans fil ont permis le développement de minuscules systèmes micro-électroniques à faible coût, appelés capteurs, capable de détecter et de réagir par rapport à un phénomène physique [1]. Ces capteurs sont souvent caractérisés par de faibles ressources telles que l'énergie, la capacité de calcul, la mémoire, etc. Un capteur a trois fonctions principales : i) la capture de données liées à un environnement physique –température, pression, vibration, lumière, mouvement, etc. ii) le traitement des données collectées ; iii) et la transmission de ces données vers un centre de traitement.

Les réseaux de capteurs sans fil avec leur champ d'application très diversifié dans les domaines militaire [4], industriel, domestique [7, 8], environnemental et de la santé [2, 19], leurs ont permis de gagner une grande popularité aussi bien dans la communauté scientifique qu'industrielle [1, 6, 12]. Les réseaux de capteurs sans fil (RCSF) sont généralement dépourvus d'une infrastructure et déployés de manière aléatoire à large échelle (ex : largage de milliers de capteurs par avion) se heurtent à de nombreux problèmes liés à la consommation énergétique, la durée de vie globale du réseau, la couverture des zones d'intérêt, l'agrégation des données, l'auto-organisation, etc. Beaucoup de solutions ont été proposées dans la littérature, néanmoins le problème de la durée de vie et de la couverture nécessitent des améliorations.

Notre contribution dans ce document concerne un mécanisme de maximisation de la durée de vie du réseau basé sur la gestion des états des nœuds capteurs. Nous introduisons dans ce papier le concept de nœud sentinelle à travers un mécanisme que nous dénommons *SENTINEL*. *SENTINEL* se propose dans un premier temps d'exploiter la densité du réseau (beaucoup de redondance) en permettant l'activation d'un sous ensemble minimum (les nœuds sentinelles) pour assurer la surveillance de la zone d'intérêt. Les nœuds redondants sont ainsi mis en veille afin d'économiser leur batterie. Ces derniers doivent se réveiller de temps à autre pour éventuellement prendre le relais lorsqu'un nœud sentinelle est «mort». Les réveils sont gérés grâce à la distribution de probabilité de *Weibull* afin d'ajouter plus de robustesse au mécanisme.

Le reste de ce document est organisé comme suit : la section 2, présente une revue de la littérature. La section 3, présente une vue d'ensemble de notre proposition, *SENTINEL*. Ensuite nous discutons à la section 4 de l'implémentation et des expérimentations. Enfin nous terminons à la section 5 par une synthèse et les perspectives pour nos travaux futurs.

2. État de l'art

La réduction de la consommation d'énergie et la maximisation de la durée de vie des capteurs occupent aujourd'hui une place prépondérante dans le domaine de la recherche scientifique dans les réseaux de capteurs sans fil [3, 9, 10, 13, 15, 17]. Cependant, malgré les nombreux mécanismes, algorithmes et protocoles proposés au niveau du routage, de l'agrégation de données, de la mobilité, de la sécurité et surtout de la couverture des zones d'intérêt, et de la conservation de l'énergie ; beaucoup parmi eux nécessitent d'être améliorés. Virmani et al. se proposent de maximiser la durée de vie avec un modèle d'agrégation des données basé sur le *clustering*[17]. Leurs travaux s'appuient sur la minimisation

de la distance de communication inter-nœuds. Dans cette même optique, Murthy et al. proposent dans [13] un protocole de clustering qui s'exécute au niveau MAC. La plupart des travaux sur la maximisation de la durée de vie des réseaux de capteurs sans fil se propose une gestion efficace et efficace du problème la couverture. Jeni et al. proposent dans [9] un algorithme distribué garantissant la couverture avec une durée de vie d'un facteur de $O(\log n * \log n B)$ où B représente l'énergie initiale d'un capteur. Dans [5, 16, 18, 21], les auteurs utilisent la répartition de la zone d'intérêt en plusieurs sous-ensembles (disjoints et/ou non disjoints) afin de rationaliser l'utilisation de l'énergie.

D'autres travaux sont axés sur des mécanismes et protocoles de gestion distribuée des états des capteurs et ceux de la couverture de la zone d'intérêt. Cela suppose que les nœuds coopèrent entre eux pour prendre des décisions par rapport aux sous ensembles à activer ; d'où la nécessité de synchroniser l'activité des nœuds [14, 20]. Cette approche exige un certain coût de traitement et de communication au niveau de chaque nœud du réseau. Cependant, elle est plus «scalable» et plus souple par rapport aux défaillances des nœuds. Dans PEAS [20], Ye et al. proposent un mécanisme qui s'appuie sur les états des nœuds capteurs et procèdent à une activation partielle des nœuds et la mise en veille des autres redondants. Cette solution permet d'économiser la consommation de l'énergie et est dotée d'un mécanisme de réactivation efficace mais ne prend pas en compte l'usure des nœuds et les pannes imminentes.

3. SENTINEL

Dans cette section, nous proposons *SENTINEL*, un mécanisme de gestion efficace des états des nœuds capteurs tout en prenant en compte le problème de la couverture de la zone d'intérêt. L'algorithme *SENTINEL* est une extension de *PEAS* et est composé de deux principales phases : une phase de sondage afin de vérifier l'état des voisins suivie d'une phase de réactivation. Nous considérons un réseau dense avec des nœuds capteurs déployés de manière aléatoire et initialement à l'état de veille.

3.1. Description des états et fonctionnement d'un nœud sentinelle

Nous considérons qu'un nœud sentinelle peut se mettre sur l'un des quatre états suivants : veille, sondage, actif ou mort. Un nœud sentinelle est initialement en mode *veille*, la radio est éteinte, et de ce fait ne communique pas avec ses voisins. Le nœud initialement en veille passe au bout d'un certain temps en mode **sondage** (Figure-1,arc-1) afin de détecter dans son voisinage un nœud en activité «*qui a monté la garde*». S'il détecte un nœud sentinelle, il retourne en mode veille (Figure-1,arc-2) sinon il «monte la garde» c'est-à-dire passe en mode **actif** (Figure-1,arc-3) jusqu'à épuiser sa réserve énergétique, passage à l'état «*dead*» (Figure-1,arc-5), ou qu'il retourne en mode *veille* (Figure-1, arc-4) à la réception d'un *probe replay* d'un autre sentinelle plus âgé (les nœuds sentinelles gardent à jours leur âge d'activité afin de pouvoir désactiver les plus jeunes). Et le modèle de gestion des états dans *SENTINEL* permet une commutation entre ces différents états. Nous avons introduit la phase **actif-veille** pour résoudre le problème de l'activation due aux collisions.

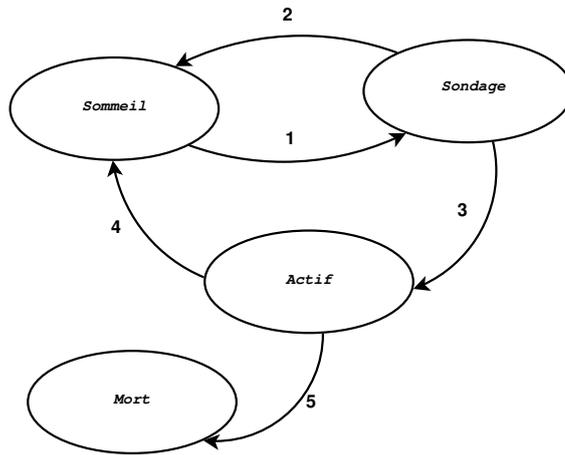


Figure 1. États d'un nœud sentinelle

3.2. Gestion efficace de l'énergie

La gestion efficace et efficace de l'énergie passe par une activation partielle des nœuds répartis à travers la zone d'intérêt. Le modèle *SENTINEL*, comme beaucoup de solutions existantes, s'appuie sur la redondance des nœuds capteurs afin de prolonger la durée de vie du réseau. Après déploiement, les nœuds ont l'autonomie de contrôler de façon distribuée l'activation d'un sous ensemble minimal de capteurs pour assurer le « monitoring » et de mettre la partie redondante en mode veille afin de conserver l'énergie. Les nœuds en veille, se réveillent après un temps t_s pour sonder leur voisinage. Le temps de veille des capteurs redondants suit une distribution de weibull (Équation 1) permettant de prendre en compte l'usure en fonction du temps mais aussi une mesure dynamique du taux de sondage des nœuds sentinelles. Un nœud sentinelle, initialement en mode veille pendant un temps t_s , se réveille afin de détecter la présence d'un nœud sentinelle en envoyant des messages de sondage puis déclenche un temporisateur. Les sentinelles qui recevront les messages, vont répondre pour signaler leur présence. À la réception d'une réponse, le nœud « sonde » calcule un nouveau temps de veille puis retourne en mode veille. Cependant, s'il ne reçoit aucune réponse de ses voisins jusqu'à l'expiration de son temporisateur t_w , il considère qu'aucun nœud n'a monté la garde et entre aussitôt en activité.

$$f(t) = \begin{cases} \frac{\beta(t-\gamma)^{\beta-1}}{\alpha^\beta} e^{-\left(\frac{t-\gamma}{\alpha}\right)^\beta} & t, \alpha, \beta > 0 \\ 0 & \text{sinon} \end{cases} \quad (1)$$

Où α est le paramètre d'échelle, β le paramètre de forme et γ le paramètre de position. À $t = 0$ on a $\gamma = 0$ et l'équation 1 est réduite à :

$$f(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta e^{-\left(\frac{t}{\alpha}\right)^\beta}} \text{ avec } t, \alpha, \beta > 0 \quad (2)$$

Pour déterminer les temps de veille des nœuds, nous posons R la probabilité de réveil d'un nœud donné avec :

$$R = 1 - F(t) \quad [3]$$

$$= 1 - P[X \leq t] \quad [4]$$

$$= 1 - \left(\int_0^t f(u) du \right) \quad [5]$$

$$= 1 - \left(1 - e^{-\left(\frac{t}{\alpha}\right)^\beta} \right) \quad [6]$$

$$= e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad [7]$$

Ainsi, à partir de la relation [7], nous en déduisons le temps de veille t .

$$\ln R = - \left(\frac{t}{\alpha} \right)^\beta \quad (8a)$$

$$\ln \frac{1}{R} = \left(\frac{t}{\alpha} \right)^\beta \quad (8b)$$

$$\frac{t}{\alpha} = \left(\ln \left(\frac{1}{R} \right) \right)^{\frac{1}{\beta}} \quad (8c)$$

$$t = \alpha \ln \left(\frac{1}{R} \right)^{\frac{1}{\beta}} \quad (8d)$$

Un autre apport significatif avec le mécanisme *SENTINEL*, est que les nœuds ont la possibilité d'ajuster dynamiquement leur taux de sondage λ , avec $\alpha = \frac{1}{\lambda}$, grâce à la fonction de hasard $h(t)$. Nous posons :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t < X \leq t + \Delta t \mid X > t) \quad (9a)$$

$$= \frac{f(t)}{1 - F(t)} \quad (9b)$$

$$= \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} \quad (9c)$$

Ainsi, à la réception d'une réponse ou «*probe reply*» d'un nœud sentinelle, le nœud «sonde» réajuste son taux de sondage (relation-9c) puis calcule son nouveau temps de veille. Pour ce qui est du nœud sentinelle, dès le début de son activité, il déclenche un temporisateur t_{act} qui mesure le temps d'activité. Ce temps sera inclus dans les messages de réponses aux nœuds sondes. Étant donné que les réponses sont envoyées en diffusion dans le voisinage, d'autres sentinelles peuvent les recevoir et dans ce cas, ils vont comparer leur durée d'activité à celle contenue dans la réponse reçue ; si leur propre durée d'activité est supérieure à celle du voisin, ils ignorent la réponse et continuent leur activité, sinon ils réajustent leur taux de sondage, les nœuds calculent à nouveau un temps de veille puis retournent en mode veille afin de laisser le nœud sentinelle voisin assurer le monitoring. L'introduction de cette étape nous a permis de régler les problèmes engendrés par les collisions. Du fait qu'il y ait une possibilité que des collisions se produisent, un

nœud peut attendre jusqu'à l'expiration de son temporisateur t_w sans recevoir de réponses de ses voisins et de ce fait commence son activité.

4. Implémentation et expérimentations

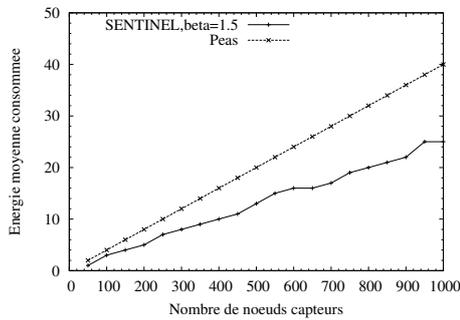


Figure 2. Energie moyenne consommée pour $\beta = 1,5$

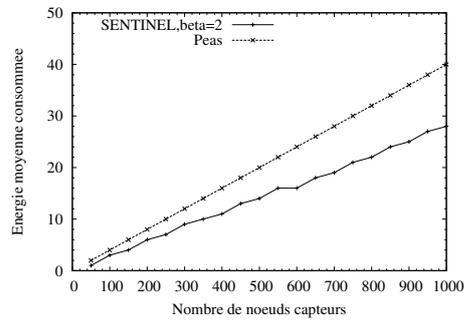


Figure 3. Energie moyenne consommée pour $\beta = 2$

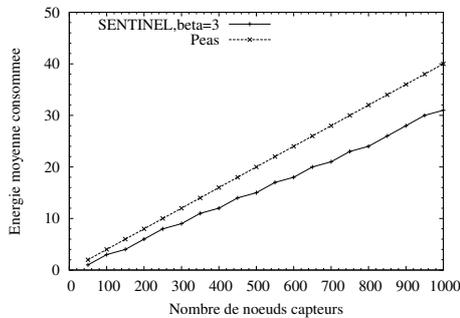


Figure 4. Energie moyenne consommée pour $\beta = 3$

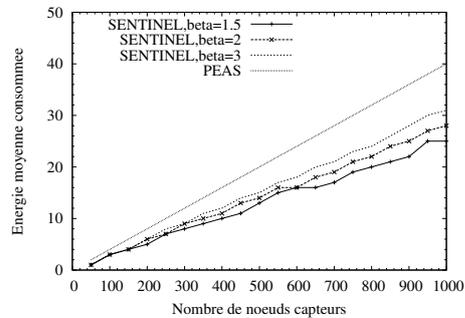


Figure 5. Energie moyenne consommée pour différentes valeurs de β

Pour l'expérimentation du mécanisme *SENTINEL*, nous avons utilisé Castalia¹, un framework d'OMNeT++² dédié aux réseaux de capteurs sans fil. Nous considérons dans nos études expérimentales, un réseau uniformément dense de $50m \times 50m$ avec les paramètres d'initialisation, $\lambda_{init} = 0,012$ et la taille du réseau entre de 50 à 1000. Nous avons fait tourner les simulations pendant un temps $t_{simul} = 600secondes$ avec des valeurs de $\beta = 1,5, 2, 3$. Les résultats des simulation montrent une nette réduction de la consommation énergétique du réseau par notre mécanisme. *SENTINEL* s'appuie sur la distribution de probabilité de *Weibull* qui présente une certaine particularité ; c'est-à-dire permet de retrouver les lois de probabilité *Exponentielle* si $\beta = 1$ et *Rayleigh* si $\beta = 2$. Le choix porté sur la distribution de *Weibull* se justifie par le fait qu'il permet

1. <http://castalia.research.nicta.com.au/index.php/en/>
 2. www.omnetpp.org

de prendre en compte l'usure des nœuds en fonction du temps. De ce fait, les temps de veille sont ajustés en fonction de l'âge du réseau. Et pour différentes valeurs de β , les expérimentations montrent que le mécanisme *SENTINEL* gère plus efficacement la consommation énergétique par rapport à la distribution choisie et par rapport à *PEAS* avec un gain de 35,40427%. Parallèlement, *SENTINEL* permet d'avoir un réseau avec plus de longévité comparé à *PEAS*.

5. Conclusion

Ce travail traite du problème de la maximisation de la durée de vie dans les réseaux de capteurs sans fil en utilisant une technique de gestion des états des nœuds capteurs. Nous avons proposé dans le cadre de ce papier *SENTINEL*, un mécanisme de maximisation de la durée de vie du réseau s'appuyant sur la gestion des états des nœuds capteurs pour la minimisation de la consommation énergétique. *SENTINEL* est un modèle utilisant la distribution de probabilité de *WEIBULL* et permet à travers les résultats de réduire la quantité moyenne d'énergie consommée par l'ensemble des nœuds du réseau d'un taux de 35% par rapport à *PEAS*. Ce qui nous permet de conclure qu'il assure une maximisation de la durée de vie du réseau.

Nous envisageons dans nos travaux futurs d'améliorer davantage les performances mais aussi de l'étendre pour apporter des éléments de solutions aux problèmes de la couverture et de la mobilité.

6. Bibliographie

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks : a survey. *Computer networks*, 38(4) :393–422, 2002.
- [2] H. Alemdar and C. Ersoy. Wireless sensor networks for healthcare : A survey. *Computer Networks*, 54(15) :2688–2710, 2010.
- [3] O. Bilenne, S.A. Attia, and J. Raisch. Distributed lifetime maximisation in wireless sensor networks using cooperation techniques. In *Communications, Computing and Control Applications (CCCA), 2011 International Conference on*, pages 1–6, march 2011.
- [4] T. Bokareva, W. Hu, S. Kanhere, B. Ristic, N. Gordon, T. Bessell, M. Rutten, and S. Jha. Wireless sensor networks for battlefield surveillance. In *Proceedings of the land warfare conference*, 2006.
- [5] M. Cardei, M. Thai, Y. Li, and W. Wu. Energy-Efficient target coverage in wireless sensor networks. *IEEE*, March 2005.
- [6] A. D de Mazieux, V. Gauthier, M. Marot, J. Vaudour, and M. Becker. *Rapport de recherche INT No 05001 RST*, 2006.
- [7] J.K. Hart and K. Martinez. Environmental sensor networks : A revolution in the earth system science ? *Earth-Science Reviews*, 78(3) :177–191, 2006.
- [8] S. Hussain, S. Schaffner, and D. Moseychuck. Applications of wireless sensor networks and rfid in a smart home environment. In *Communication Networks and Services Research Conference, 2009. CNSR'09. Seventh Annual*, pages 153–157. IEEE, 2009.
- [9] J.J. Jeni, PG Student, D.S. Senthil, and G.K. Sundararaj. Lifetime and coverage guarantees through distributed life time maximization and distributed sensor cover algorithms in wireless

- sensor networks. 2012.
- [10] V. Katiyar, N. Chand, and S. Soni. Lifetime maximization of wireless sensor networks through energy map : A review. *Wireless Sensor Network*, 2011.
 - [11] L. Khelladi and N. Badache. 2004.
 - [12] N. Mitton. AUTO-ORGANISATION DES RESEAUX SANS FIL MULTI-SAUTS a GRANDE ECHELLE. 2006.
 - [13] J.K. Murthy, S. Kumar, and A. Srinivas. Energy efficient scheduling in cross layer optimized clustered wireless sensor networks. *Int'l Journal of Computer Science and Communication*, 3(1) :149–153, 2012.
 - [14] S. Sengupta, S. Das, M. Nasir, A.V. Vasilakos, and W. Pedrycz. An evolutionary multiobjective sleep-scheduling scheme for differentiated coverage in wireless sensor networks. *Systems, Man, and Cybernetics, Part C : Applications and Reviews, IEEE Transactions on*, 42(6) :1093–1102, 2012.
 - [15] G. Shirazi and L. Lampe. Lifetime maximization in uwb sensor networks for event detection. *Signal Processing, IEEE Transactions on*, 59(9) :4411–4423, sept. 2011.
 - [16] M. T Thai, F. Wang, D. H Du, and X. Jia. Coverage problems in wireless sensor networks : designs and analysis. *International Journal of Sensor Networks*, 3(3) :191–200, 2008.
 - [17] D. Virmani and S. Jain. Decentralized lifetime maximizing tree with clustering for data delivery in wireless sensor networks. *arXiv preprint arXiv :1201.4943*, 2012.
 - [18] J. Wu, M. Cardei, F. Dai, and S. Yang. Extended dominating set and its applications in ad hoc networks using cooperative communication. *IEEE Transactions on Parallel and Distributed Systems*, pages 851–864, 2006.
 - [19] H. Yan, Y. Xu, and M. Gidlund. Experimental e-health applications in wireless sensor networks. In *Communications and Mobile Computing, 2009. CMC'09. WRI International Conference on*, volume 1, pages 563–567. IEEE, 2009.
 - [20] F. Ye, G. Zhong, J. Cheng, S. Lu, and L. Zhang. PEAS : a robust energy conserving protocol for long-lived sensor networks. In *Distributed Computing Systems, 2003. Proceedings. 23rd International Conference on*, pages 28–37. IEEE, 2003.
 - [21] D. Zorbas, D. Glynos, P. Kotzanikolaou, and C. Douligeris. Solving coverage problems in wireless sensor networks using cover sets. *Ad Hoc Networks*, 8(4) :400–415, 2010.

Softcust

A Friendly Localized Platform for Multilingual Semantic Communication

Fabian Cretton, Zhan Liu and Anne Le Calvé

Institute of Business Information Systems
University of Applied Sciences of Western Switzerland
TechnoArk 3
3960 Sierre
SWITZERLAND
fabian.cretton@hevs.ch

RÉSUMÉ. Les dernières tendances de l'économie et de la mondialisation ont créés de nouveaux besoins pour améliorer la communication et la coopération dans les échanges internationaux. Cependant, il est souvent difficile de découvrir des informations pertinentes entre différentes langues et cultures dans les systèmes de gestion de l'information conventionnels. Dans ce papier, nous proposons un outil de communication sémantique et multilingue pour une gestion d'information automatisée dans un système convivial et localisé. Le modèle théorique est composé d'une ontologie multilingue basée sur Wikipédia pour fournir un moteur de recherche et un système de suggestion puissants. Notre solution actuelle comprend l'anglais, le français et le chinois.

ABSTRACT. The new economy and the globalization trends have brought new needs for increasing international business communication and cooperation. However, it is often difficult to discover relevant information across languages and cultures in conventional information management system. In this study, we have proposed a multilingual semantic communication for automated information management through a user-friendly localized system. The theoretical model is made of a multilingual ontology based on linguistic knowledge from Wikipedia in order to provide a powerful search engine and suggestion system. Our current solution includes English, French and Chinese.

MOTS-CLÉS : communication sémantique multilingue, gestion des données et des connaissances, ontologie multilingue, moteur de recherche sémantique, plate-forme de business international.

KEYWORDS: multilingual semantic communication, knowledge and data management, multilingual ontology, semantic search engine, international business platform.

Actes du 5e Colloque National sur la Recherche en Informatique et ses Applications –
Ziguinchor, Sénégal, Avril 2013

1. Introduction

Given the strong demand of the Chinese markets for professional management software, we have conducted applied research on localizing Swiss-made computer programs on the Chinese market during the Sino-Swiss Software Customization Network Project¹. This research included several technical topics such as intellectual property protection, Chinese characters display or ePayment. This previous project was successful in localizing a few products for the Chinese market, but some repetitive research tasks could be improved and partly automated, especially for matching the Swiss-made products with the existing demands on the Chinese market. Therefore we present this SoftCust project with the main goal to develop an intelligent matching platform to help the Chinese and Swiss IT SMEs match their competences, needs, offers and joint activities intelligently.

As stated by the W3C, the Semantic Web «provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries»². Nowadays semantic web knowledge's representation (ontologies) is a popularized mean of designing and sharing a domain model and its data. This is true about any domain (e.g. health care, finance or business), and also about representing a language's lexical semantic as well as cross-language information. Moreover, semantic tools provide powerful ways for men and machines to process data, for instance using specific features of the query language or using rules engines (i.e. reasoners) to infer new knowledge and perform data validation checks.

Therefore our research question is: how can semantic web technologies improve the performance of an exchange platform in designing a friendly localized system with multilingual semantic communication.

In the rest of the paper we address our research question using a design science approach and we structure our paper according to [12] guidelines: the first part discusses some related work for the semantic matching aspect. In the second part we present the chosen solution and then its implementation in the third part. Some discussion and evaluation is presented in the fourth part. We then conclude with some future directions, improvements and challenges to overcome in the fifth part.

2. Related Work

To achieve semantic matching in a multilingual exchange platform, we mainly

¹ <http://iet.hevs.ch/valais/softcust-sino-suisse-software-customization.html>

² <http://www.w3.org/RDF/FAQ>

tackled two topics: text mining and cross-language information retrieval.

Text mining is a domain of natural language processing (NLP), a computer science field where machines extract knowledge from text written in natural language. In most of NLP tasks machine learning tools have been successfully used for years. Downsides of that approach are the amount of human work to annotate existing corpus in order to train the system and the limitation of the system that becomes efficient only for the domain it has been trained on. Recently, knowledge based techniques are giving comparatively good results, and even better results in some cases, as demonstrated by [11] for word-sense disambiguation or [4] for text classification. In this approach, the knowledge is described in a structured manner, for instance using ontologies, and the NLP algorithms exploit that knowledge to perform their tasks on the texts.

Ontologies are becoming a common way to describe knowledge. They are based on W3C standards as RDF and OWL, or the SKOS³ vocabulary to define concept schemes, glossaries and terminologies. NLP tools making use of ontologies are spreading, especially for named entity recognition. Those tools can thus recognize entities (e.g. persons, organizations or locations) with very good results as presented in [8][13], sometimes as online services⁴. However, the core of our problematic relies more on dealing with common words (synonyms for instance) than named entities.

Information retrieval (IR) techniques have been elaborated to identify the most relevant elements of a corpus to satisfy a search. Full-text search, providing a simple match between a search key-word and that word in the indexed texts, is quite limited when it comes to lexical semantics. It can be enhanced using query expansion⁵ techniques which, as stated by [2], are becoming quite mature but still face difficulties. The same conclusions can be drawn for cross-lingual methods (CLIR), a field which produces good results but still considered a non-trivial task [14][5]. Traditional NLP and IR techniques both have to deal with lexical semantic and cross-lingual issues, for which semantic web technologies provide very interesting features. Our view of the semantic information retrieval issue is similar to the proposals of KIM [13] [17] and TAP [18], which focus on automatic population and annotation of documents. However, our solution includes Chinese and is more specifically tailored to the case.

We need to handle full-text search enhanced with semantics and cross-lingual capabilities for 3 specific languages: French, Chinese and English. This requires an ontology that is generic enough to contain the three language's lexical semantic such as synonymy, hyponymy or holonymy, about terms specific to the computer science domain, but that could be easily extended to any domain. A number of generic

³ <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>

⁴ <http://viewer.opencalais.com/>, <http://www.alchemyapi.com/>

⁵ http://en.wikipedia.org/wiki/Query_expansion

ontologies are available as OpenCyc⁶, UMBEL⁷, or Wikipedia based as DBPedia [1] or Yago [15][9]. Unfortunately, even though those resources do contain multilingual information, the languages we need are not well described and also lack lexical semantic. Other ontologies as WordNet do contain rich linguistic information, but there is currently no multilingual version of WordNet that includes French, English and Chinese in a freely available resource.

As we can see, many solutions propose the features related to our platform, but there is no integrated solution that can carry out all the tasks for French, English and Chinese. We thus needed to develop new tools and ontologies to answer our research question.

3. Solution

We designed a platform that helps users with different roles (programmer, distributor, vendor, etc.) to find a match (a user looking for a software, a vendor looking for a distributor, etc.) based on some structured data and also descriptions in natural language. The platform is a hybrid system based on a relational database and a triple store to manage classical and semantic knowledge. The semantic search is also a compound engine based on semantic indexing of text with an ontology, coupled with a classical full-text search.

A generic multilingual ontology has been created from the English, French and Chinese Wikipedia. NLP functionalities are used to analyze users' descriptions and link them with the ontology terms. Then, cross-lingual matches are performed between the requests and the offers. A main component of our platform, called «semText», allows semantic indexing of texts and then performs matches based on cross-lingual and lexical semantic defined in the ontology. We did generate a SKOS ontology from Wikipedia dumps, using the Java-based Wikipedia Library (JWPL) as presented in [16]. That solution has the advantage to be easily extended to over 200 languages found in Wikipedia, and it can also be enriched by other related information found in Wiktionary⁸ or in the LOD⁹ for instance. The resulting multilingual ontology does contain the needed information about a word: its translation in the 3 languages, synonyms, and exploitable links to other concepts (e.g. hyponyms, related words).

To handle text, NLP systems have to carry out pre-processing tasks [10] as tokenization, stemming/lemmatization, part-of-speech tagging, for which a number of tools are available. We did choose the renowned GATE architecture [3] available as a

⁶ <http://sw.opencyc.org/>

⁷ <http://www.umbel.org/>

⁸ <http://www.wiktionary.org/>

⁹ <http://linkeddata.org/>

user interface and a java API. Having the ontology on one side and the text descriptions on the other, semText uses GATE to generate a semantic lookup graph of the texts, i.e. link the words of a text to the concepts of the ontology. To describe the lookup graph we designed the OntoManagNLP ontology. Here is some information about the different components we used for GATE. The LKB (Large Knowledge Base) Gazetteer from OntoText handles the very big French and English gazetteers, representing millions of concepts. The classical GATE's ANNIE gazetteer was used for about 200'000 Chinese concepts, the LKB being not able to handle Chinese words. Tokenization being particularly subtle for the Chinese characters, we added the ICTClas¹⁰ library to both, Lucene and Gate, as the default tools were giving poor results. Finally, Tree-tagger¹¹ is used for the French part-of-speech.

The texts are thus indexed with the ontology, but also with Apache Lucene¹² for full-text search. Using the ontology's links between words to search for the texts, allows for more precise querying than using a traditional query expansion technique. When a user looks for a word, the ontology is used to look for texts which have lookups to any related word (e.g. synonym, hyponym), in a very precise manner, moreover handling directly the cross-lingual search. The need for full-text search was that no matter how generic the ontology is, it will never contain all words of a language. However, words in the indexed texts but not in the ontology must still be usable for searches even though there is no lookup for them. We thus created a hybrid search: semantic search for the keywords (or related words) found in the ontology, and full-text search for the others (handling translation on the fly for multi-lingual results).

Based on the relevant literatures, we create an artifact in the form of a model [7] to represent our multilingual semantic search engine. We adopt a design science research methodology and refer to existing guidelines for design theories [6]. In order to better understand our solution, we present a simple scenario. Suppose that Alice is a manager in a small CRM software company in Switzerland. Because of the financial crises and the increasing competition in domestic market, she is looking overseas for potential clients. Ming is running his textile company in China and is willing to buy CRM software to help him understanding the needs of its customers. However, he notices that CRM software is less mature in China and wants localized CRM software from outside of China. Unfortunately, there are some problems to get Alice and Ming connected, and that's where our solution comes into play. When Ming uses this platform with keyword search «client management software» in Chinese, he can have a list of «CRM» software companies and their product description. Therefore, our solution provides cross-lingual matching enhanced with related words handling (synonyms, generalization and specialization) in the context of a specific search.

¹⁰ <http://code.google.com/p/openictclas/>

¹¹ <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

¹² <http://lucene.apache.org/core/>

4. Implémentation

In the Intelligent Matching Platform users create their, describe their domain of activities and upload documents to add information on their products. The Softcust system manages classic multilingual information with a MySQL database. Information is additionally indexed semantically and available for search.

Partners' searches are based on several criteria: keywords, roles, domain types, countries and languages. The system performs the semantic matching and displays the results in the user's mother tongue (translation being done with the Google Translate API¹³). Searches can be saved as wish lists, and thus managed by a notification system that will notify the users whenever new corresponding entries enter the system.

The back-end triple store is OWLim 4.3 from OntoText¹⁴, chosen because it supports loading/querying/reasoning on a large amount of triples, as well as very useful SPARQL 1.1 features.

5. Discussions and Conclusions

An evaluation has been done by the SoftCust team with real data coming from their former project. They did appreciate that the underlying semantic tool works transparently on the existing texts, without having to enter new data. Results are thus very encouraging as the tool does allow performing automatic match between data in different languages and using different vocabulary or expressions.

In this research project we realized a semantic enriched platform to help Swiss and Chinese actors of the computer science field to find matches between supply and demands. For that system based on semantic web technologies we've generated a SKOS ontology from Wikipedia that includes lexical semantic of the French, English and Chinese languages. Our research question was thus answered as the semantic engine allows richer search results than simple full-text search, handling synonyms and related words (as hyponyms) in a more efficient way than traditional IR query expansion, and last but not least, with cross-lingual capabilities. For the scientific community, we provide new alternatives for international communications and business models into the management of international "buyer-seller" relationship.

In a future work, we would like to improve the evaluation of the results by implementing some IR techniques, and enrich the ontology in the different languages, adding information from the new versions of Wikipedia and complement it with

¹³ <https://developers.google.com/translate/>

¹⁴ <http://www.ontotext.com/>

Wiktionary or any data source from the LOD. SemText will be used to work on the lookup graph and implement technics of word sense disambiguation as well as text classification. Moreover, the main remark about further tests and downsides evaluation is that the system needs to contain a large amount of descriptions and searches to see how it performs in a real-time environment.

We thank Ontotext who supported us with a research license for OWLIM.

6. Bibliography and biography

6.1 Bibliography

- [1] Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R. and Ives Z., *DBpedia: a nucleus for a web of open data*. In Aberer et al. (Eds.): The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea. Lecture Notes in Computer Science 4825 Springer 2007, ISBN 978-3-540-76297-3, 2007
- [2] Carpineto C. and Romano G., *A survey of automatic query expansion in information retrieval*. ACM Computing Surveys, 44(1), 2012
- [3] Cunningham H., *GATE, a General Architecture for Text Engineering*. Computers and the Humanities, 2002
- [4] De Maat E., Krabben K., and Winkels R., *Machine Learning versus Knowledge Based Classification of Legal Texts*. In Proceedings of the 2010 conference on Legal Knowledge and Information Systems: JURIX 2010: The Twenty-Third Annual Conference, Radboud G. F. Winkels (Ed.). IOS Press, Amsterdam, The Netherlands, 87-96, 2010
- [5] Dolamic L. and Savoy J., *Monolingual and Bilingual Searches: Evaluation, Challenges and Failure Analysis*. Submitted, 2008
- [6] Gregor S. and Jones D., *The anatomy of a design theory*. Journal of the Association for Information Systems, 8(5), pp. 312-335, 2007
- [7] March S. T. and Smith G. F., *Design and natural science research on information technology*. Decision Support Systems, 15(4), pp. 251-266, 1995
- [8] Mendes P. N., Jakob M., García-Silva A., and Bizer C., *DBpedia spotlight: shedding light on the web of documents*. In Proceedings of the 7th International Conference on Semantic Systems (I-Semantics '11), ACM, New York, NY, USA, 1-8, 2011
- [9] Müller C. and Gurevych I., *Using Wikipedia and Wiktionary in domain-specific formation retrieval*. In Proceedings of the 9th Cross-language evaluation forum conference on Evaluating systems for multilingual and multimodal information access (CLEF'08), 219-226, 2008
- [10] Nadkarni P., Ohno-Machado L., Chapman W. W., *Natural language processing: an introduction*. Journal of the American Medical Informatics Association, 18: 544-551, 2011

Cretton et al.

- [11] Navigli R., *Word Sense Disambiguation: A Survey*. ACM Computing Surveys, 41(2), ACM Press, pp. 1-69, 2009
- [12] Peffers K., Tuunanen T., Rothenberger M. A., and Chatterjee S., *A Design Science Research Methodology for Information Systems Research*. Journal of Management Information Systems, 24(3), 45-77, 2007
- [13] Popov B., Kirayakov A., Ognyanoff D., Manov D., Kirilov A., *KIM - a semantic platform for information extraction and retrieval*. Natural Language Engineering 10 (3/4), pp. 375-392, 2004
- [14] Savoy, J., *Comparative study of monolingual and multilingual search models for use with Asian languages*. ACM TALIP, vol. 4(2), pp. 163–189, 2005
- [15] Suchanek F. M., Kasneci G., Weikum G., *Yago: A large ontology from Wikipedia and wordnet*. Journal of Web Semantics 6 (3) 203-217, 2008
- [16] Zesch T., Müller C. and Gurevych I., *Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary*. In Proceedings of the Conference on Language Resources and Evaluation LREC, electronic proceedings, 2008
- [17] Kiryakov A., Popov B., Terziev I., Manov D., and Ognyanoff D., *Semantic Annotation, Indexing, and Retrieval*. J. Web Semantics, vol. 2, no. 1, pp. 49-79, 2004.
- [18] Guha R. and McCool R., *TAP:a Semantic Web platform*. Computer Networks, 42(5), pp. 557–577, 2003.

6.2 Biography

The authors work at the University of Applied Sciences of Western Switzerland.

Fabian Cretton is an active analyst and senior developer specialized in semantic web technologies in various domains such as user modeling, adaptive interfaces, ontology design, knowledge management and natural language processing.

Zhan Liu is also PhD candidate at the Department of Information Systems of the University of Lausanne, Switzerland. His main research interests are semantic web, privacy and security management, mobile business and mobile technology development. His work has been published in journals, book chapters and conferences.

Dr. Anne Le Calvé is professor since 1999. She holds a PhD in computer science - about meta search engine methodologies in information retrieval domain. Her research activities focus on knowledge management, information modeling and semantic web technologies, applied to different fields such as eTourism, eGov, personal information management and user modeling.

CNRIA'2013

Vers un système de veille épidémiologique fondé sur une ontologie

Application à la bilharziose au Sénégal

Gaoussou CAMARA^{1,2,3} – Sylvie DESPRES³ – Rim DJEDIDI³ – Moussa LO^{1,2}

¹LIRIMA, Equipe M2EIPS IRD UMI 209 UMMISCO,

²LANI, UFR SAT, Université Gaston Berger, BP 234 Saint-Louis, Sénégal

³LIM&BIO, Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny

gaoussou.camara@ugb.edu.sn, sylvie.despres@univ-paris13.fr, rim.jedidi@univ-paris13.fr,
moussa.lo@ugb.edu.sn

.....
RÉSUMÉ. Les systèmes de veille épidémiologique sont utilisés pour contrôler l'évolution de la propagation des maladies en proposant des plans d'action pour prévenir les risques identifiés. Dans ce domaine, la prédiction du risque est fondée sur des approches quantitatives qui ne sont guère utilisables lorsque la collecte de données est difficile. Dans ce papier, une approche qualitative fondée sur une ontologie de la veille épidémiologique est proposée. Nous décrivons la conception de cette ontologie. Puis, nous montrons comment une ontologie s'intègre dans les systèmes de surveillance classiques et permet de pallier les limites liées aux approches quantitatives. L'approche adoptée est appliquée au contexte de la veille de la bilharziose au Sénégal en impliquant les experts du domaine.

ABSTRACT. Epidemiological monitoring systems are used to control the evolution of disease spreading and to suggest action plans to prevent identified risks. In this domain, risk prediction is based on quantitative approaches that are hardly usable when data collection is tricky. In this paper, a qualitative approach based on an epidemiological monitoring ontology is proposed. We describe the design of this ontology and show how it fits into classical monitoring systems and helps overcoming limits related to quantitative approaches. This approach is exemplified through the monitoring of the schistosomiasis in Senegal by integrated local experts.

MOTS-CLÉS : Veille épidémiologique, Analyse de risque, Prédiction précoce, Simulation, Modélisation ontologique, Bilharziose.

KEYWORDS: Epidemiological monitoring, Risk analysis, Early prediction, Simulation, Ontological modeling, Schistosomiasis.

.....

1. Introduction

Les systèmes de veille épidémiologique sont utilisés pour contrôler la propagation des maladies. Nous proposons une approche qualitative de la veille fondée sur une ontologie pour améliorer cette situation. L'usage de cette ontologie dans le domaine de la veille permettra : (i) la réalisation des simulations qualitatives [6] pouvant se substituer aux simulations numériques lorsque les données ou les modèles numériques nécessaires sont indisponibles ; (ii) la spécification des connaissances associées aux modèles de simulation numériques pour faciliter leur composition ou leur interopérabilité ; (iii) l'accès à une terminologie facilitant la communication et le travail collaboratif entre les acteurs des différentes organisations impliquées dans la veille.

Dans ce papier, nous présentons la conception d'une ontologie pour la veille épidémiologique. Puis, nous montrons comment elle s'intègre dans les systèmes de veille épidémiologique classiques pour pallier les limites propres aux approches quantitatives. Le papier est organisé comme suit. Dans la section 2, nous présentons le processus de la veille épidémiologique et les limites liées à sa mise en œuvre. Dans la section 3, nous détaillons l'approche ontologique proposée. Nous illustrons en section 4 notre contribution avec le cas de la veille de la bilharziose [10] au Sénégal. Dans la dernière section, nous concluons et présentons nos travaux futurs.

2. La veille épidémiologique

2.1. Principe de la veille

La veille épidémiologique est une activité consistant à surveiller de façon continue la propagation d'une maladie dans une population située dans une région géographique donnée afin d'anticiper et contrôler son évolution. Elle permet d'identifier des facteurs de risque. Lorsqu'un événement lié à ces facteurs est détecté, toutes les données nécessaires à l'analyse de son impact dans la propagation de la maladie sont collectées. L'analyse de ces données consiste à réaliser des simulations numériques afin de prédire précisément le risque sur la propagation lié à l'occurrence de cet événement. La méthode d'analyse mise en œuvre dans ces systèmes dépend de la nature du phénomène surveillé. Ces types de phénomènes, par leurs évolutions et leurs émergences issues des interactions des éléments les composant, sont qualifiés de systèmes complexes. Pour analyser, afin de prédire, l'évolution et la dynamique de tels systèmes, il est indispensable de recourir à une modélisation systémique [9]. Plusieurs de ces approches ont montré leur intérêt pour l'étude des systèmes complexes comme les approches mathématiques et les approches informatiques fondées sur les systèmes multi-agents [1], [5]. La simulation de ces phénomènes requiert la disponibilité de modèles de simulation et de données pour répondre aux questions susceptibles de se poser à propos de la propagation de la maladie.

2.2. Limites dans la mise en œuvre du processus de veille

La collecte des données constitue une phase incontournable de l'analyse des événements détectés par le système de veille. Cette collecte automatique doit se faire en temps réel pour garantir à la fois une simulation et une estimation rapide et précise du risque. Cependant, la mise en place d'un tel dispositif requiert des moyens techniques qui ne sont pas toujours disponibles.

Le dispositif de veille doit également favoriser un partage de connaissances et d'expériences entre les différentes structures impliquées dans la veille. En outre, les acteurs participant à l'analyse des risques et à la prise de décision peuvent avoir des profils hétérogènes (épidémiologiste, biostatisticien, pathologiste, météorologue, médecin, entomologiste, parasitologue, acteur de santé publique, acteur politique, etc.), et par conséquent des vocabulaires différents pour désigner les mêmes concepts du domaine. Il est par conséquent nécessaire de trouver un consensus sur un vocabulaire commun pour faciliter leur communication, leur interaction et leur travail collaboratif.

3. Un système de veille épidémiologique à base d'ontologie

3.1. Objectif

Notre objectif est de proposer une approche alternative et complémentaire permettant aux systèmes de veille de répondre à leurs fonctions dans toutes les situations possibles. Elle doit permettre de : (i) s'affranchir des contraintes liées à l'acquisition de données numériques pour effectuer une simulation et une prédiction rapide des risques potentiels ; (ii) expliquer l'origine ou la cause d'une observation ; (iii) faciliter, au besoin, l'interopérabilité permettant la composition des modèles de simulation numériques existants ; (iv) faciliter le partage de connaissances entre les organismes intervenant dans le domaine de la veille ; (v) faciliter la communication et le travail collaboratif entre les différents acteurs du domaine.

Pour relever ces défis, l'approche proposée est fondée sur la construction d'une ontologie de la veille épidémiologique. Une telle ontologie permet de répondre aux différentes situations décrites puisqu'elle fournit une description qualitative des processus permettant de modéliser le comportement du phénomène sans paramètre numérique. Cette spécification permet une reproduction du processus de propagation d'une maladie à partir de la description abstraite de ses processus internes. Elle définit également les différents états inhérents aux entités impliquées dans le processus de propagation. Ainsi, une simulation qualitative utilisant des modèles ontologiques spécifiant les processus permettra de prédire des états ou des occurrences d'événements possibles liés à la maladie sans données numériques. L'ontologie définit un vocabulaire commun facilitant la communication entre les acteurs impliqués dans l'utilisation du système et favorisant un meilleur partage des connaissances entre les organismes

intervenant dans le domaine et une meilleure interopérabilité entre les modèles de simulation numériques. Elle permet également une représentation formelle permettant une simulation qualitative sur ces descriptions afin d'effectuer des prédictions rapides, d'aider à l'interprétation de phénomènes observés ou des résultats des simulations numériques et d'assister la mise en œuvre de plans d'action.

3.2. Modélisation ontologique des connaissances

La nécessité de modéliser conjointement les connaissances du domaine, les connaissances sur les processus et leurs relations est une caractéristique essentielle de la veille épidémiologique. Les concepts de processus, d'événement, d'état et d'objet et les relations les liant correspondant à un haut niveau d'abstraction sont modélisés dans les ontologies dites de fondement [7]. Notre ontologie du domaine de la veille épidémiologique est construite en réutilisant une ontologie noyau [3] s'appuyant sur une ontologie de fondement. L'ontologie de la veille est organisée en deux modules correspondant respectivement aux connaissances du domaine et à celles sur les processus. Les relations entre ces deux modules sont également représentées dans l'ontologie.

3.3. Architecture d'un système de veille à base d'ontologie

L'architecture de notre système [4] est fondée sur une ontologie de la veille épidémiologique et un ensemble de modèles de simulation respectivement utilisés par le module d'assistance et le simulateur numérique pour répondre aux requêtes d'analyse. La surveillance permet de détecter des événements liés aux facteurs de risque de la maladie. Le module d'assistance récupère les requêtes d'analyse. Les requêtes sont traitées par simulation numérique si les modèles et les données sont disponibles. Dans le cas contraire un raisonnement sur les modèles ontologiques des processus du domaine permet une simulation qualitative. Si les modèles et leurs paramètres sont disponibles, le module d'assistance supervise la mise en œuvre des simulations en assistant la sélection et la composition des modèles numériques. Une alerte est lancée en cas de risques de propagation de la maladie. Des décisions sont ensuite élaborées et évaluées par une approche quantitative ou qualitative pour anticiper sur les risques identifiés.

4. Cadre d'application : veille de la bilharziose au Sénégal

Cette section décrit l'apport de la modélisation des connaissances sur le domaine et celles sur des processus *via* quatre cas d'utilisation sur la bilharziose au Sénégal : la prédiction ; la composition de modèles de simulation numériques ; l'assistance à l'explication d'observations et l'assistance à la mise en œuvre des plans d'action.

4.1. La bilharziose au Sénégal

La bilharziose est une maladie parasitaire constituant un problème de santé publique important au Sénégal et il n'existe pas de système informatique capable de répondre de manière automatique et rapide à ces besoins. La lenteur au niveau de la collecte des données retarde la prédiction du risque. En outre, il n'existe pratiquement pas de modèle de simulation numérique utilisable pour faire des prédictions. C'est pourquoi le cas de la bilharziose au Sénégal se prête à une approche qualitative de simulation pour la veille épidémiologique.

4.2. Périmètre de l'ontologie de la veille de la bilharziose

Afin de définir le périmètre de l'ontologie, nous avons élaboré avec les experts du programme national de lutte contre la bilharziose (PNLB) le scénario intitulé « Risques liés à l'eau ». Il nous permet de définir quatre cas d'utilisation possibles de l'ontologie au cours des différentes phases du processus de veille épidémiologique.

4.2.1. Scénario : Risques liés à l'eau

L'eau est à la fois le vecteur de la transmission de la bilharziose et l'environnement de multiplication des parasites. Il est donc nécessaire de surveiller les sources potentielles pouvant favoriser la présence de points d'eau ou les activités pouvant entraîner le contact des hôtes avec les points d'eau. Dans cette étude de cas, la surveillance porte sur les pluies.

4.2.2. Prédiction précoce en l'absence de données

En ayant connaissance de l'occurrence de l'événement pluie, une simulation par raisonnement qualitatif sur l'ontologie permettra de prédire son impact sur la prévalence de la maladie dans le futur. Des mesures provisoires pourront alors être prises pour anticiper sur les risques possibles, en attendant de disposer de données plus fines comme la quantité de pluie tombée.

4.2.3. Assistance à la simulation numérique

Si, par les moyens de collecte, nous obtenons la quantité de pluie tombée et toutes les données nécessaires pour prédire de façon précise l'impact sur la prévalence, alors nous pouvons procéder à la simulation numérique. Le module d'assistance peut, grâce à l'ontologie, identifier les modèles de simulation et demander à l'utilisateur de les fournir. Les résultats de la simulation sont également envoyés au module d'assistance pour, éventuellement, un formatage pour la visualisation des résultats, l'extraction et la capitalisation de connaissances produites au cours de la simulation, l'assistance à l'interprétation des résultats par l'utilisateur, etc.

4.2.4. Assistance à l'explication d'une observation

Le raisonnement sur l'ontologie des processus peut aider à l'interprétation de l'origine d'une observation en s'appuyant sur les préconditions et les relations de précédence entre les processus. Il permet ainsi d'identifier les causes de l'émergence ou de la propagation de la maladie. Par exemple, l'occurrence d'une nouvelle infection est le résultat de l'occurrence d'une contamination car la contamination est précédée par la phase d'exposition dans le modèle SEIRS [3] de la bilharziose. Il existe donc au moins un point d'eau source de contamination dans la zone ou ses environs. Par conséquent, diverses décisions peuvent être prises selon que l'hôte infecté (i) vit dans la même localité où il a été diagnostiqué (bloquer l'accès aux points d'eau le temps d'identifier la source de contamination et de procéder à la désinfection) ou (ii) vient d'ailleurs (dans ce cas, lancer une alerte aux autorités sanitaires de la localité d'où il vient).

4.2.5. Assistance à la mise en œuvre des plans d'action

Un raisonnement qualitatif sur les processus, notamment les processus liés aux actions possibles associées aux risques, permettra l'élaboration de plans d'action et la sélection du plus approprié. Ces actions peuvent être modélisées comme des processus dont les conditions d'occurrence et les effets sont modélisés dans l'ontologie. Ainsi, il est possible de contrôler l'exécution du plan d'action et assister sa mise en œuvre, d'en donner les étapes suivantes après l'exécution d'une action et la vérification de ses effets.

4.3. Une ontologie pour la veille de la bilharziose

Le contenu d'une ontologie est défini par ses objectifs. Dans le contexte de la veille épidémiologique, nos deux objectifs majeurs sont : (i) faciliter le travail collaboratif entre les différents acteurs, l'échange de données entre les organisations et la réalisation des simulations numériques ; (ii) planifier et superviser des processus de la veille et la réalisation de simulation qualitative comme solution alternative aux contraintes liées à l'acquisition des données quantitatives ou la disponibilité de modèles numériques. Pour répondre à ces besoins, il est donc nécessaire de disposer : (i) d'une ontologie de domaine qui fournit un vocabulaire, une représentation des connaissances du domaine et une description des modèles de simulation ; (ii) d'une ontologie des processus de la veille et de la propagation de la maladie.

4.3.1. Les connaissances du domaine

Les connaissances du domaine portent sur la bilharziose, ses facteurs de risque et les modèles de simulation numériques existants. La bilharziose est causée par un agent pathogène du genre des *Trématodes*. Ce genre appartient à la famille des *Schistosomatidae*. Selon les espèces de mollusque vivant au Sénégal, seules deux espèces de parasite y survivent : *S. Haematobium* et *S. Mansoni* causant respectivement

la bilharziose urinaire et la bilharziose intestinale. Les *facteurs de risque* regroupent l'ensemble des éléments pouvant avoir une influence sur la propagation de la maladie.

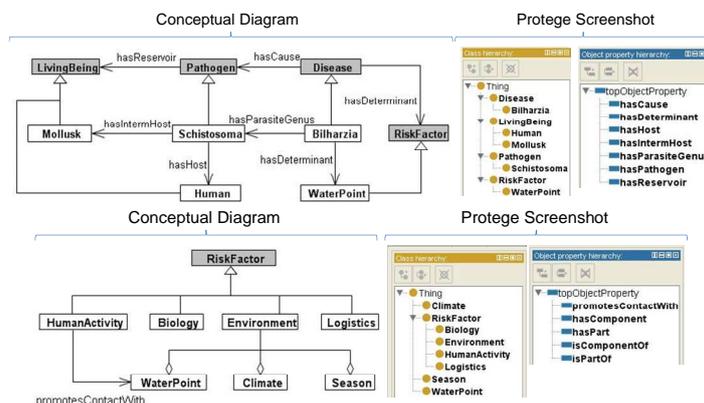


Figure 1– Modèles conceptuels : bilharziose et facteurs de risque

Les modèles numériques permettent d'analyser, par des simulations, les processus liés à la bilharziose. Ils reproduisent une partie de la dynamique de la bilharziose avec des objectifs précis tels que la prédiction de la prévalence. Ils sont décrits dans l'ontologie avec leurs fonctions objectives, leurs entrées et leurs sorties.

4.3.2. Les processus du domaine de la veille épidémiologique

La veille épidémiologique est modélisée selon deux macro-processus encapsulés : le *processus de veille* et le *processus de propagation* du phénomène épidémiologique [4]. Le *processus de veille* est composé de : (i) la surveillance comportant les processus de détection des événements à risque et de collecte des données pour l'analyse des événements à risque ; (ii) l'analyse ; (iii) la prise de décisions ; (iv) l'évaluation des décisions. Dans cette section, nous présentons uniquement la modélisation du macro-processus de propagation de la bilharziose au sein de la population. Ce modèle fait apparaître les liens existants entre les états de l'hôte et les processus transitoires qui leur sont associés.

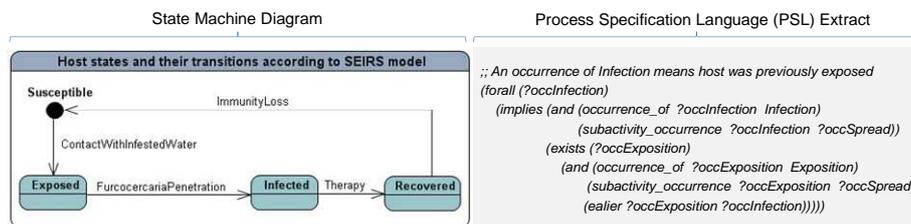


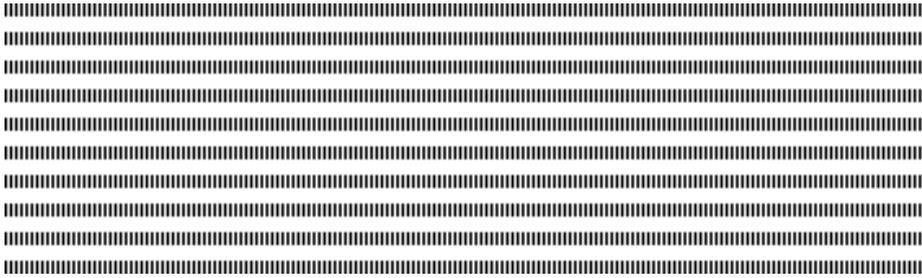
Figure 2 – Modèle SEIRS de propagation de la bilharziose

5. Conclusion

Ce papier présente une nouvelle approche fondée sur les ontologies pour la mise en œuvre du processus de la veille épidémiologique. L'ontologie modélise à la fois les connaissances du domaine et les processus sous-jacents afin de faciliter leur partage, la communication entre les acteurs de la veille, la réalisation des prédictions rapides par des simulations qualitatives et l'élaboration et la mise en œuvre des décisions. L'apport d'une telle approche est argumenté dans le cadre de la veille de la schistosomiase au Sénégal en montrant comment l'ontologie pourrait être utilisée tout au long du processus de veille. Néanmoins, la modélisation ontologique de la dynamique des systèmes complexes reste très superficielle. Ainsi dans le cas de la schistosomiase, les échelles temporelles et spatiales doivent être prises en compte. En outre, si la simulation qualitative donne l'ensemble des possibilités des risques, elle ne précise pas leur probabilité d'apparition. Il convient par conséquent de prendre des décisions optimisées intégrant cette incertitude mais limitant également les risques.

6. Bibliographie

- [1] Amouroux, E., Taillandier, P. and Drogoul, A. (2010) Complex environment representation in epidemiology ABM: application on H5N1 propagation. *In the 3rd ICTACS*. Can Tho City, Vietnam.
- [2] Buton, F. (2006) De l'expertise scientifique à l'intelligence épidémiologique : l'activité de veille sanitaire, *Genèses*, 4, 65,71-91.
- [3] Camara, G., Despres, S., Djedidi, R., and Lo, M. (2012) Vers une ontologie des processus de propagation des maladies infectieuses. *In Actes IC 2012*, Paris.
- [4] Camara G, Despres S, Djedidi R, and Lo M. Towards an ontology for an Epidemiological Monitoring System, Proc. of the 9th International ISCRAM Conference – Rothkrantz L, Ristvej J and Franco Z eds., Vancouver, 2012.
- [5] Drogoul, A., Treuil, J.-P. and Zucker, J.-D. (2008) Modélisation et simulation à base d'agents : Exemples commentés, outils informatiques et questions théoriques, Dunod, Paris.
- [6] Forbus, K. D. (1997) Qualitative Reasoning, *The Computer Science and Engineering Handbook*, 715-733.
- [7] Grenon P, Smith B, and Goldberg L. Biodynamic Ontology: Applying BFO in the Biomedical Domain, In: *Stud. Health Technol. Inform.* IOS Press, 2004; pp. 20–38.
- [8] Gruninger, M. (2009) Using the PSL Ontology, *In S. Staab, Handbook of Ontologies*, 419-431, Springer-Verlag, Berlin.
- [9] Le Moigne, J.-L. (1990) La modélisation des systèmes complexes, Dunod, Paris.
- [10] Talla, I., Kongs, A. and Verlé, P. (1992) Preliminary study of the prevalence of human schistosomiasis in Richard-Toll (the Senegal river basin), *Trans R Soc Trop Med Hyg*, 86, 2, 182-191.



Gestion des services dans P2P4GS

Gueye Bassirou*# — Flauzac Olivier* — Rabat Cyril* — Niang Ibrahima#

* CReSTIC, UFR Sciences Exactes et Naturelles
Université de Reims Champagne Ardenne
FRANCE
bassirou.gueye@etudiant.univ-reims.fr, {olivier.flauzac, cyril.rabat}@univ-reims.fr

Département de Mathématiques et d'Informatique
Université Cheikh Anta Diop de Dakar
SENEGAL
bassirou.gueye@ucad.edu.sn, iniang@ucad.sn

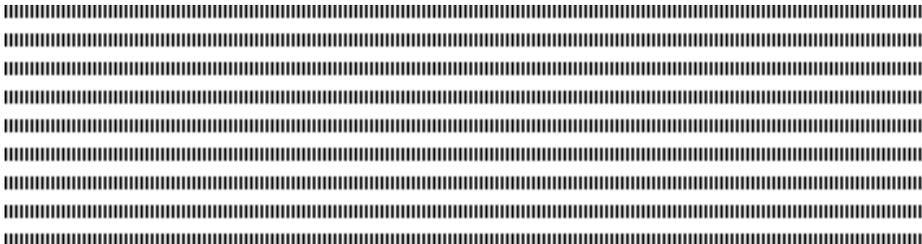


RÉSUMÉ. Les grilles qui exploitent la notion de services sont basées sur des architectures hiérarchiques fortement centralisées. Cette centralisation implique une gestion unifiée des ressources, mais aussi des difficultés à réagir vis à vis des pannes et des fautes qui impactent la communauté. Ainsi, nous avons proposé dans nos travaux précédents une spécification originale de gestion de services dans un environnement de grilles de calcul basé sur le paradigme pair à pair (P2P). Cette spécification nommée P2P4GS offre un ensemble de primitives (« deploy », « lookup », « invoke », « exec », « save ») dont pour chacune d'entre elles, différentes stratégies ou modèles de fonctionnement ont été proposés. Dans ce papier, nous proposons une étude de complexités algorithmiques des primitives de déploiement et de localisation de services dans P2P4GS en les plongeant sur les topologies classiques de la pile P2P à savoir l'anneau et l'arbre.

ABSTRACT. The grids that use the concept of services are based on highly centralized hierarchical architectures. The main issue of this centralization is the unified management of resources and the difficult to react promptly against failures and faults that can affect grid users. Thus, we have proposed in our previous work an original specification of services management in a grid computing environment based on peer-to-peer (P2P) paradigm. This specification named P2P4GS provides a set of primitives (« deploy », « lookup », « invoke », « exec », « save ») whose for each of them, different operating strategies have been proposed. In this paper, we propose a study of algorithmic complexities of deployment and service localization primitives in P2P4GS by immersing them on the classical topologies of P2P stack ie the ring and tree.

MOTS-CLÉS : Réseaux pair à pair, Grilles de calcul, Services web

KEYWORDS : Peer to peer network, Grid computing, Web services.



1. Introduction

Les grilles de calcul (Grid Computing) [6, 5, 8, 14] sont une technologie en pleine expansion dont le but est d'offrir aux organisations virtuelles, ainsi qu'à la communauté scientifique des ressources informatiques virtuellement illimitées. L'apparition des Services Web [11] a fourni un cadre qui a initié l'alignement de ces deux technologies et qui a été à l'origine des grilles de services [9, 10]. Ces grilles de services ont permis de distribuer les traitements, pour obtenir une utilisation optimale des ressources. Un effort particulier de normalisation est actuellement fait afin d'apporter, comme dans le cas des services web, toutes les ressources et les moyens nécessaires au développement d'applications dans ces grilles de services. Cependant, ces grilles sont basées sur des architectures hiérarchiques fortement centralisées [5, 14]. Cette centralisation implique une gestion unifiée des ressources, mais aussi des difficultés à réagir vis à vis des pannes et des fautes qui impactent la communauté.

Cet article propose une étude de complexités algorithmiques des primitives de déploiement et de localisation de services définies dans [1, 2, 3]. En fait, la spécification nommée *P2P4GS* [3] que nous avons proposée présente l'originalité de ne pas lier l'infrastructure pair à pair à la plate-forme d'exécution de services. Elle se voudra la plus générique possible, non liée à un réseau P2P particulier ou à un protocole de gestion de services défini à l'avance. Il serait par conséquent très utile voire fondamental d'étudier l'impact de ces primitives sur les topologies classiques existantes à savoir l'anneau et l'arbre.

La suite de ce document est organisée comme suit. Dans un premier temps, nous allons étudier les travaux connexes, puis présenter l'originalité de notre solution. Ensuite, nous présenterons le modèle que nous utilisons, ainsi que les différents aspects de notre spécification. Par la suite, nous étudierons la complexité des primitives « *deploy* » et « *lookup* » de la spécification *P2P4GS*. Nous terminerons cet article par une conclusion, ainsi que des perspectives futures.

2. Travaux connexes

L'exécution de code distant a été définie de différentes façons. L'appel distant de code a donné lieu au schéma général des RPC (*Remote Procedure Call*). La spécification *RPC* spécifie les échanges entre un consommateur d'informations (client) et la ressource (serveur). Différentes implémentations qui mettent en œuvre différents langages, protocoles ou systèmes ont été proposées, dont celles basées : (i) sur des bibliothèques systèmes : *ONC RPC* ; (ii) sur des objets à distance : *JAVA RMI*, *Corba* ; (iii) sur l'utilisation de composants, *J2EE*, *Net - Mono* ; (iv) sur les services web [11] : *XML-RPC*, *SOAP*.

Le concept de *RPC* a été transféré dans les grilles, soit à partir des bibliothèques permettant la mise en place de solutions techniques, comme *Globus* [6], soit à partir des solutions directement conçues pour assurer le *GRID RPC* [13, 14]. Les solutions proposées sont hiérarchiques et nécessitent des points de centralisation de la connaissance. C'est le cas par exemple de *Ninf-G* [13] et *DIET* [14] dans lesquelles les capacités d'exécution s'enregistrent auprès d'un point de centralisation.

Comme dans le cas des exécutions sur le Web, des solutions d'implantation de services dans les grilles [10, 9] et de découverte de ressources dans de tels environnements

ont été proposées. Cependant, l'ensemble de ces propositions est basé sur des solutions d'architectures hiérarchiques qui présentent un degré de dynamique très réduit et dont les services d'infrastructure, comme le déploiement et la localisation de services, sont eux même centralisés.

Dans le cadre d'architecture pair à pair, on note une gestion décentralisée de la topologie. C'est le cas par exemple de Pastry [7] qui propose une organisation en anneau. Par conséquent, les spécifications de la localisation des ressources ainsi que de la gestion de la topologie sont liées : une seule spécification est donnée pour ces deux charges. D'autre part, une architecture hiérarchique de réseau pair à pair basée sur le modèle existant de découverte de ressources dans une grille a été proposée dans [16].

3. Contribution

La spécification P2P4GS [3] que nous avons proposée présente l'originalité de ne pas lier l'infrastructure P2P à la plate-forme d'exécution de services. En effet, contrairement à des solutions de type Pastry [7] nous proposons de séparer la couche de gestion de la grille P2P des couches de localisation et d'exécution de services. La spécification est générique et tend donc à être applicable sur toute architecture pair à pair.

Les performances que nous avons obtenues d'une manière générale sont très satisfaisantes tant sur la topologie anneau orienté ou non orienté que sur l'arbre. En effet, avec une performance de $O(n)$ au pire des cas, la topologie anneau orienté s'adapte mieux à nos algorithmes en terme de nombre de messages échangés et celle de l'anneau non orienté s'adaptera mieux à ces algorithmes en terme de temps d'exécutions. Soulignons que plus le diamètre de l'arbre est faible plus nos algorithmes deviennent meilleur avec cette topologie.

4. Vue d'ensemble de notre spécification P2P4GS [3]

Afin de définir notre spécification, nous décrivons notre modèle d'application basé sur 3 composants : le réseau, les nœuds qui composent le réseau et la notion de service.

Le réseau que nous considérons est le réseau overlay mis à notre disposition par la plate-forme pair à pair. Afin d'être exploitable dans notre spécification, le réseau est modélisé sous forme d'un graphe $G = (V ; E)$ où V est l'ensemble de nœuds et E l'ensemble des liens de communications.

Les nœuds sont en charge de la gestion locale du réseau, ils assurent le réceptacle des services, c'est à dire la plate-forme d'exécution. Chaque nœud détient une table des services qui répertorie l'ensemble des services hébergés et ceux dont il a connaissance de leurs localisations.

Les services sont des objets intégrables aux plates-formes d'exécution détenues en chaque nœud. Un service est caractérisé par sa plate-forme d'exécution, les ressources nécessaires à son exécution (CPU, RAM, données, etc.), le format et contenu nécessaires lors de son invocation et les contraintes des données résultats.

Nous définissons un modèle d'architecture en couche (Figure 1) dans laquelle nous allons intégrer notre middleware. Cette modularisation nous permettra d'hériter des ser-

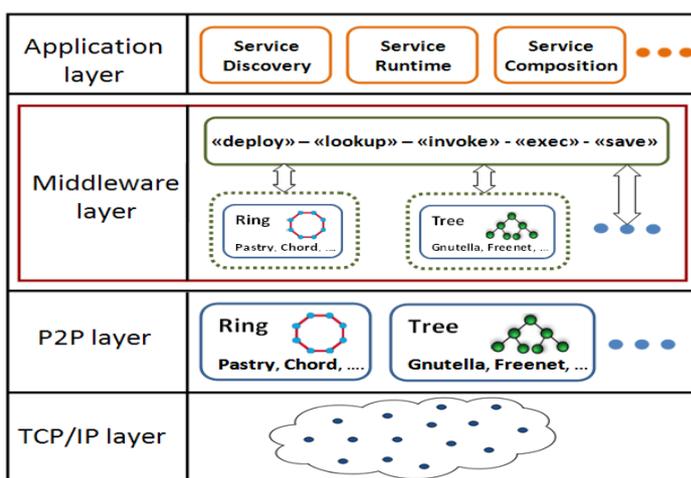


Figure 1 – Architecture de la spécification P2P4GS

vices offerts par les couches basses. Ainsi, la couche TCP/IP qui représente la couche physique pourrait offrir des fonctions ou services de sécurité, de fiabilité, etc [12].

Le couche P2P est une couche logique qui exploite un protocole de réseau pair à pair (Pastry, Gnutella, etc.) reposant sur une topologie (anneau, arbre, etc.).

La couche middleware représente notre spécification P2P4GS [3]. Elle est constituée de deux sous couches. Une sous couche directement liée à la topologie P2P sous-jacente et qui permet de faire abstraction de cette dernière. Une sous couche qui propose un ensemble de primitives (« deploy », « lookup », « invoke », « exec », « save ») [1, 2, 3] offrant toutes les solutions de gestion de services qui vont de la création ou implantation d'un service jusqu'à sa consommation. Ces primitives sont exploitées par la sous couche d'abstraction afin d'offrir des services (service runtime, service composition, etc.) à la couche application.

5. P2P4GS : Etude de complexités algorithmiques des primitives de déploiement et de localisation de services

Dans cette section, nous présentons une étude de la complexité des primitives « deploy » et « lookup » de la spécification *P2P4GS*. Pour ce faire, nous décrivons d'abord leurs principes d'exécution (les algorithmes ainsi que les détails de leurs fonctionnements sont disponibles dans [4]). Ensuite, nous mesurons les performances de ces algorithmes en fonction des topologies classiques de la pile P2P à savoir l'anneau et l'arbre.

5.1. Déploiement d'un service dans P2P4GS

Nous décrivons ici la stratégie de déploiement équilibrée. Pour les stratégies "aléatoire" et "premier nœud" [1, 2, 3], nous les étudierons en simulations. La stratégie de déploiement équilibrée consiste à collecter les statistiques des différents nœuds et à tenter d'équilibrer la charge en terme de nombre de services hébergés entre ces nœuds.

Pour déployer un service dans P2P4GS, il suffit de se connecter à un nœud dit point d'entrée qui se chargera d'initier le processus d'élection du nœud sur lequel sera déployé le nouveau service. Précisons que tout nœud dans notre l'architecture peut faire office de point d'entrée.

Principe d'exécution : il faut trouver le nœud capable d'héberger le nouveau service et ayant un plus petit nombre de services. Les étapes suivantes sont prévues :

1) Le nœud qui initie le processus doit être à l'état disponible (*isAvailable()*). En effet, si le nœud choisi comme point d'entrée n'est pas disponible, cela signifie qu'un processus d'élection sollicité par un autre utilisateur est en cours d'exécution.

2) Le nœud vérifie s'il est apte (*isAble()*, c'est à dire s'il respecte les contraintes en terme de CPU, RAM, plateforme d'exécution, etc.) et s'il détient un plus petit nombre de services ; au quel cas il sera candidat pour l'hébergement du nouveau service.

3) Le nœud initiateur envoie la requête *queryDeploy()* à :

- son voisin (*next*) s'il s'agit d'une topologie anneau orienté ;
- ses deux voisins (*next* et *prev*) pour une topologie anneau non orienté ;
- tout ses voisins (*neigh*) sauf à son père, dans le cas d'une topologie arbre.

4) Tout nœud qui reçoit la requête *queryDeploy()* effectue l'opération 2.

5) L'opération 3 est effectuée par :

- cas de l'anneau orienté : tout nœud ayant reçu la requête *queryDeploy()* sauf le nœud initiateur ;

- cas de l'anneau non orienté : tout nœud ayant reçu la requête *queryDeploy()* au plus une fois. Si un nœud reçoit une deuxième fois une requête *queryDeploy()* (surement de son autre voisin *next* (respectivement *prev*)), il envoie une requête *resultDeploy()* à son voisin *prev* (respectivement *next*) ;

- cas de l'arbre : tout nœud ayant reçu la requête *queryDeploy()* et ayant au moins 2 voisins. Si un nœud n'a qu'un seul voisin (son père), il lui répond par un message *resultDeploy()* après avoir effectué l'opération 2.

6) Le processus d'élection se termine à la suite d'une décision du nœud initiateur :

a) s'il existe au moins un nœud capable d'héberger, celui qui disposera le plus petit nombre de services sera élu ;

b) sinon, le déploiement sera impossible.

7) Evénement déclenchant un processus de décision :

- cas de l'anneau orienté : la réception d'un message *queryDeploy()* par le nœud initiateur ;

- cas de l'anneau non orienté : la réception d'un message *resultDeploy()* par le nœud initiateur. Si un message *resultDeploy()* est reçu par un nœud autre que l'initiateur, ce dernier le transmet à son suivant ;

- cas de l'arbre : la réception par le nœud initiateur d'un message *resultDeploy()* de tous ses voisins. Si un message *resultDeploy()* est reçu par un nœud autre que l'initiateur, ce dernier effectue l'opération 2. Lorsque tous les voisins auxquels il avait envoyé une requête *queryDeploy()* lui répondent, il transmet le résultat à son père par un *resultDeploy()*.

8) Après exécution du processus d'élection, le nœud initiateur se remet à l'état disponible.

5.2. Localisation d'un service dans P2P4GS

Pour localiser un service dans P2P4GS, le premier nœud sollicité sera le point d'entrée et ce dernier se chargera d'initier le processus de localisation du service.

Principe d'exécution :

1) Le nœud initiateur vérifie d'abord s'il a déjà initié le même processus de localisation ; au quel cas il ne traitera pas la nouvelle demande mais répondra juste à la fin du processus en cours. Sinon, le processus de localisation du service sera initié.

2) Recherche locale du service avec la fonction *lookup()*.

a) La recherche s'arrête si le service est présent sur le nœud auquel *lookup()* a été exécuté ou si le nœud a connaissance de la localisation du service :

- si la localisation est effectuée au niveau du nœud initiateur ou si ce dernier reçoit un *resultLookup()*, il envoie comme résultat l'adresse du nœud détenant le service ;
- sinon, le nœud ayant localisé le service envoie le résultat à son père et de façon récursive le résultat sera acheminé jusqu'au nœud initiateur (donc le chemin inverse de la requête de localisation du service).

b) S'il n'y a pas de connaissance locale sur la localisation du service, alors envoi de la requête *queryLookup()* à :

- son voisin (*next*) s'il s'agit d'une topologie anneau orienté ;
- ses deux voisins (*next* et *prev*) pour une topologie anneau non orienté ;
- tout ses voisins (*neigh*) sauf à son père, dans le cas d'une topologie arbre.

3) Tout nœud qui reçoit la requête *queryLookup()* effectue l'opération 2.

4) L'opération 2.b) est effectuée par :

- cas de l'anneau orienté : tout nœud ayant reçu un *queryLookup()* sauf l'initiateur ;
- cas de l'anneau non orienté : tout nœud ayant reçu la requête *queryLookup()* au plus une fois. Si un nœud reçoit une deuxième fois une requête *queryLookup()* (surement de son autre voisin *next* (respectivement *prev*)), il envoie une requête *resultLookup()* à son voisin *prev* (respectivement *next*) ;
- cas de l'arbre : tout nœud ayant reçu la requête *queryLookup()* et ayant au moins 2 voisins. Si un nœud n'a qu'un seul voisin (son père), il lui répond par un message *resultLookup()* après avoir effectué l'opération 2.

5) Evénement déclenchant l'arrêt du processus de localisation :

- cas de l'anneau orienté : la réception d'un *queryLookup()* par le nœud initiateur ;
- cas de l'anneau non orienté : la réception d'un message *resultLookup()* par le nœud initiateur. Si un message *resultLookup()* est reçu par un nœud autre que l'initiateur, ce dernier le transmet à son suivant ;
- cas de l'arbre : la réception par le nœud initiateur d'un message *resultLookup()* de tous ses voisins. Si un message *resultLookup()* est reçu par un nœud autre que l'initiateur, ce dernier effectue l'opération 2. Lorsque tous les voisins auxquels il avait envoyé une requête *queryLookup()* lui répondent, il transmet le résultat à son père par un *resultLookup()*.

6) A la fin du processus de localisation d'un service, le nœud initiateur enlève la requête de localisation de sa pile.

5.3. Etude de la complexité des primitives « deploy » et « lookup »

Pour évaluer les performances de nos algorithmes de déploiement et de localisation, nous utilisons comme critère de mesure le nombre de messages (M) échangés entre les nœuds de la communauté et de temps d'exécution (T). Le tableau 1 ci contre illustre les complexités des nos algorithmes en fonction des topologies classiques de la pile P2P à savoir l'anneau et l'arbre. n représente le nombre de nœuds.

Complexités	Nombre de messages (M)		Temps d'exécution (T)	
	deploy	lookup	deploy	lookup
Anneau orienté	$O(n)$	$2 \leq M \leq 2*O(n-1)$	$O(n)$	$2 \leq T \leq 2*O(n-1)$
Anneau non orienté	$2*O(n)$	$3 \leq M \leq 2*O(n)$	$O(n)$	$2 \leq T \leq O(n)$
Arbre	$2*O(n-1)$	$2 \leq M \leq 2*O(n-1)$	$2 \leq T \leq 2*O(n-1)$	

Tableau 1 – Complexités des primitives «deploy» et «lookup» en fonction des topologies anneaux et arbre

D'une manière générale, nous remarquons que la topologie anneau orienté offre des performances supérieures. En effet, l'algorithme de déploiement est très efficace sur cette topologie tant en terme de messages échangés qu'en temps d'exécution. L'algorithme de localisation reste efficace si le nombre de sauts nécessaires pour atteindre le service recherché est acceptable. Cependant, plus le nombre de saut pour atteindre le service recherché devient important, plus on tend vers le pire des cas ($2*O(n - 1)$).

Nous remarquons également que la topologie anneau non orienté consomme beaucoup plus de messages tant en ce qui concerne le déploiement que la localisation. Cependant, cette topologie reste meilleure en termes de temps d'exécution ($O(n)$ au pire des cas).

Pour le cas de l'arbre, les bornes sont dues au fait qu'il peut prendre différentes formes de représentation (chaîne, arbre équilibré, étoile) et donc son diamètre peut varier de $n - 1$ à 2. Ainsi, plus le diamètre de l'arbre est faible plus nos algorithmes deviennent meilleur avec cette topologie.

6. Conclusion et perspectives

Dans cet article, nous avons présenté une étude de la complexité des algorithmes de déploiement et de localisation de services dans P2P4GS. Les performances que nous avons obtenues d'une manière générale sont très satisfaites tant sur la topologie anneau orienté ou non orienté que sur l'arbre. En effet, avec une performance de $O(n)$ au pire des cas, la topologie anneau orienté s'adapte mieux à nos algorithmes en terme de nombre de messages échangés et celle de l'anneau non orienté s'adaptera mieux à ces algorithmes en terme de temps d'exécutions. Soulignons que plus le diamètre de l'arbre est faible plus nos algorithmes deviennent meilleurs avec cette topologie.

Comme perspective immédiate, nous évaluerons les performances des nos algorithmes par le biais de simulations. Ce travail sera suivi par une implémentation ainsi quâune étude de solutions de tolérance aux pannes et aux fautes.

7. Bibliographie

- [1] B. GUEYE, O. FLAUZAC, I. NIANG, "Services pour les grilles pair à pair", *In Actes du 4e Colloque National sur la Recherche en Informatique et ses Applications (CNRIA). URED'2012.*
- [2] B. GUEYE, O. FLAUZAC, I. NIANG, "Services pour les grilles pair à pair", *Eleventh African Conference on Research in Computer Science and Applied Mathematics (CARI'12), Algeria, October 2012.*
- [3] B. GUEYE, O. FLAUZAC, I. NIANG, "P2P4GS : A Specification for Services Management in Peer-to-Peer Grids", *Third Kuwait Conference on e-Services and e-Systems (KCESS'2012), December 2012.*
- [4] B. GUEYE, O. FLAUZAC, C. RABAT, I. NIANG, "P2P4GS : étude algorithmique des primitives", *Technical Report, CReSTIC-SysCom (URCA) and LID (UCAD), Janvier 2013, [Online] : <http://cosy.univ-reims.fr/~bgueye/>*
- [5] I. FOSTER, C. KESSELMAN, S. TUECKE, "The anatomy of the grid : enabling scalable virtual organizations", *IJSA'01*, vol. 3, 2001.
- [6] I. FOSTER, "Globus Toolkit version 4 : Software for Service-Oriented Systems", *JSCT*, vol. 21 n° 4, July 2006.
- [7] A. ROWSTRON, P. DRUSCHEL, "Pastry : Scalable, distributed object location and routing for large-scale peer-to-peer systems, *In Proc. IFIP/ACM Middleware 2001, Heidelberg, Germany.*
- [8] O. FLAUZAC, M. KRAJECKI, L.A. STEFFENEL, "CONFIIT a middleware for peer-to-peer computing", *Journal of supercomputing*, vol. 53 n° 1, July 2010.
- [9] I. FOSTER, H. KISHIMOTO, A. SAVVA and all., "The Open Grid Services Architecture", *Informational Document, Global Grid Forum (GGF)*, vol. 6 n° 6, 2005.
- [10] S. TUECKE, K. CZAIKOWSKI, I. FOSTER, J. FREY, S. GRAHAM, C. KESSELMAN, D. SNELLING, P. VANDERBILT, "Open Grid Services Infrastructure", *Global Grid Forum Draft Recommendation*, 2003.
- [11] W.A. NAGUY, F. CURBERA, S. WEERAWARANA, "Web services : Why and how ?", *ACM OOPSLA, Workshop on Object-Oriented Web Services*, 2001.
- [12] O. FLAUZAC, F. NOLOT, C. RABAT, L.A. STEFFENEL, "Grid of Security : a decentralized enforcement of the network security", *in Manish Gupta, John Walp, and Raj Sharman (Eds.), Threats, Countermeasures and Advances in Applied Information Security*, April 2012.
- [13] Y. TANAKA, H. NAKADA, S. SEKIGUCHI, T. SUZUMURA, SATOSHI MATSUOKA, "Ninf-G : A Reference Implementation of RPC-based Programming Middleware for Grid Computing", *Journal of Grid Computing*, 2003.
- [14] E. CARON, F. DESPREZ, "DIET : A Scalable Toolbox to Build Network Enabled Servers on the Grid", *International Journal of High Performance Computing Applications*, vol. 20, 2006.
- [15] F. BUTT, S.S. BOKHARI, A. ABHARI, A. FERWORN, "Scalable Grid resource discovery through distributed search.", *International Journal of Distributed and Parallel Systems (IIDPS), Vol.2, No.5, September 2011.*
- [16] D. CHEN, G. CHANG, X. ZHENG, D. SUN, J. LI, X. WANG, "A Novel P2P Based Grid Resource Discovery Model", *In Proceedings of Journal Of Network*, pp. 1390-1397, 2011.

Minimisation de la Puissance totale dans les Systèmes Multi-Utilisateurs OFDM avec l'Algorithme Génétique

NDIAYE Abdourahmane — OUYA Samuel — MENDY Gervais —

FARSSI S. Mohamed

Ecole Supérieure Polytechnique de Dakar

Département Génie Informatique

Laboratoire Informatique, Réseaux et Télécoms

rahabdou2003@yahoo.fr samuel.ouya@ucad.sn gervaismendy@ucad.sn farsism@yahoo.com

.....

RÉSUMÉ. Les systèmes cellulaires 3G+ se basent sur la modulation multi-porteuse OFDM pour pallier aux variations du canal dues au phénomène multitrajet. Ce dernier confère au canal de transmission un comportement sélectif en fréquence. Ainsi la puissance requise sur une sous-porteuse dépendra du gain canal variable et spécifique à chaque utilisateur. L'allocation des fréquences aux utilisateurs doit se faire de manière à minimiser la puissance totale tout en satisfaisant les besoins des utilisateurs. Ce problème est une optimisation sous contrainte dont il s'agit dans ce document de faire la résolution en utilisant l'algorithme génétique qui par son mécanisme de fonctionnement, est très adapté pour une recherche plus efficiente de l'optimum global.

ABSTRACT. 3G+ cellular systems are based on OFDM multicarrier modulation to compensate variations of the channel caused by multipath phenomenon. This gives the transmission channel frequency selective behavior. The power required on a subcarrier depends on the gain channel and variable specific to each user. The allocation of frequencies to users should be to minimize the total power while meeting the users' needs. This problem is a constrained optimization which is in this document to the resolution using the genetic algorithm that its mechanism of operators is very suitable for a more efficient research of global optimum.

MOTS-CLÉS : OFDM, phénomènes multitrajet, allocation adaptative, algorithme génétique

KEYWORDS : OFDM, Multipath Effect, adaptative allocation, Genetic Algorithm.

.....

1. Introduction

Les systèmes cellulaires 3G+ utilisent la technique de multiplexage par division de fréquences orthogonales pour d'une part lutter contre la sélectivité fréquentielle du canal et d'autre part accroître le débit des utilisateurs. Dans les systèmes 2G et 3G, une sous-porteuse est utilisée lors d'une communication ce qui peut provoquer des pertes de signal si le canal de transmission est sélectif en fréquence. Cependant, dans la technique de transmission OFDM la largeur de bande totale disponible est divisée en plusieurs sous-bandes étroites [11]. Chaque sous-canal a approximativement un gain de canal supposé constant [12]. L'orthogonalité réduit l'écart entre les sous-porteuses, augmentant ainsi l'efficacité spectrale. L'abondance du nombre de sous-porteuses permet une gestion plus efficace en adaptant l'allocation des sous-porteuses par rapport au profil canal des utilisateurs. Cela devrait permettre une consommation d'énergie réduite.

Dans les systèmes multi-utilisateurs OFDM, où les ressources (puissances, sous-porteuses) doivent être partagées entre les utilisateurs, le problème d'allocation optimale se pose puisque le gain canal spécifique à chaque utilisateur varie en fonction de la fréquence. L'allocation adaptative des ressources est la méthode proposée pour améliorer la performance des systèmes cellulaires. Il est établi dans [2] qu'une allocation adaptative des ressources peut offrir un gain de 20 dB par rapport à une allocation non adaptative.

L'affectation des ressources des systèmes multi-utilisateurs OFDM se résume à la résolution d'un problème d'optimisation avec contraintes. L'optimisation prend en compte l'état du canal lorsque celui-ci peut être estimé. Dans la pratique, les informations relatives à l'état du canal sont fournies par le récepteur via le canal logique CQICH (Channel Quality Indicator Channel) de la station de base [5]. Notons que ce problème d'optimisation avec contraintes est NP-difficile. Par conséquent il est improbable que les algorithmes polynomiaux en temps, puissent être utilisés pour résoudre cette optimalité [6].

Ainsi plusieurs algorithmes ont été développés pour trouver une solution approchée au problème d'allocations de ressources. L'algorithme Waterfilling proposé par [3] pour la résolution du problème de minisation de la puissance totale, s'est montré optimal dans le cas mono-utilisateur. Des améliorations ont été apportées par [9] pour réduire la complexité de l'algorithme Waterfilling. L'algorithme de Greedy proposé par [13] s'est montré optimal pour l'affectation de bits dans le cas mono-utilisateur.

Cependant l'algorithme Greedy et l'algorithme Waterfilling sont itératifs et leur complexité évolue donc de manière linéaire avec le nombre d'utilisateurs et de sous-porteuses. C'est dans ce contexte que les algorithmes évolutionnaires inspirés de phénomènes naturels ont été introduits [4] en vue d'une résolution plus efficace du problème d'allocation de ressources.

2. Formulation du problème

Nous considérons un système multi-utilisateurs OFDM dont les caractéristiques du canal sont connues par l'émetteur et le récepteur. Le canal est supposé linéaire et sélectif en fréquence, avec un bruit additif blanc Gaussien de moyenne nulle. Le bande totale du canal est décomposée en N sous-canaux dont chacun est caractérisé par un gain canal constant spécifique à chaque utilisateur pour une sous-porteuse donnée.

Notons qu'une sous-porteuse est exclusivement allouée à un utilisateur et dans ce cas les interférences inter-symboles sont nulles.

Nous considérons un système monocellulaire avec un nombre d'utilisateurs et de sous-porteuses constant. Les utilisateurs sont supposés en micromobilité et l'effet DOPPLER est ignoré.

Soit $H_{k,n}$ le gain channel de l'utilisateur k sur la sous-porteuse n . La puissance requise est donnée par[8] :

$$p_{k,n} = \frac{f(c_{k,n})}{H_{k,n}^2} \quad [1]$$

Où

$$f(c_{k,n}) = \frac{N_0}{3} (2^{c_{k,n}} - 1) [Q^{-1}(\frac{BER_k}{4})]^2 \quad [2]$$

$c_{k,n}$ est le nombre de bits de l'utilisateur k sur la sous-porteuse n .

N_0 est la densité spectrale de puissance du bruit additif Gaussien .

BER_k est le Taux d'Erreur Binaire de l'utilisateur k .

$Q(x) = \text{erfc}(x)$ est la fonction d'erreur complémentaire.

La puissance totale allouée aux utilisateurs est donnée par :

$$P_T = \sum_{k=1}^K \sum_{n=1}^N p_{k,n} \cdot \rho_{k,n} \quad [3]$$

La quantité totale de bits affectée à l'utilisateur k est donnée par :

$$r_k = \sum_{n=1}^N c_{k,n} \cdot \rho_{k,n} \quad [4]$$

Où $\rho_{k,n} = 1$ si la sous-porteuse n est allouée à l'utilisateur k et $\rho_{k,n} = 0$ sinon.(nous considérons que la sous-porteuse n est allouée exclusivement à l'utilisateur k)

Le problème d'allocation de ressources peut s'écrire sous la forme :

$$\min(P_T) \quad [5]$$

sous-contrainte :

$$r_k > r_0 \quad [6]$$

Où r_0 est le débit minimal de l'utilisateur k .

3. Algorithme Génétique

L'algorithme génétique est inspiré de la biologie notamment sur la théorie de DARWIN et sur les travaux de MENDEL relatifs à la recombinaison des espèces. Il a été introduit pour la résolution de problèmes d'optimisation. Le principe de l'algorithme génétique est basé sur l'évolution d'une population initiale sous l'effet d'opérateurs tels que : sélection, croisement, mutation[7]. Au terme du processus d'évolution le meilleur individu de la population suivant un critère bien défini, constituera la solution au problème d'allocation de ressources.

Les différentes phases de l'algorithme génétique sont :

- Codage des chromosomes

Un chromosome ou individu de la population représente un schéma d'allocation de ressources. Le codage du chromosome consiste à donner une structure correspondant à l'allocation de ressources. Dans notre implémentation un chromosome est représenté par un tableau de structure contenant un nombre fixe de sous-porteuses numérotées de 1 à N . Chaque cellule renferme le numéro de l'utilisateur à qui la sous-porteuse courante (de la cellule) est allouée.

Structure d'un chromosome

Sous-porteuse	1	2	---	N
Utilisateur	i	j	---	k

- Initialisation de la population

L'initialisation de la population consiste à générer aléatoirement tous les chromosomes de la population. Dans notre travail la taille de la population est fixe et n'évolue pas au cours du processus.

- Evaluation :

Cet opérateur permet le calcul de la puissance totale requise et le débit totale pour chaque chromosome.

- Selection :

Après l'évaluation des chromosomes, la sélection consiste à retenir les meilleurs chromosomes soit directement par tri (sélection naturelle) ou par tournoi entre deux chromosomes sélectionnés arbitrairement (sélection par tournoi). Il existe d'autres types de sélection. Le meilleur chromosome ou individu de la population est celui qui requière moins de puissance tout en satisfaisant la contrainte de débit.

- Mutation :

La mutation consiste à introduire des changements à un certain nombre de chromosomes

pris arbitrairement dans le but d'avoir une exploration plus complète de l'espace de recherche. Dans notre travail le nombre et les positions de mutation d'un chromosome sont déterminés aléatoirement.

– Croisement :

Il consiste à créer un nouveau chromosome fils à partir de deux chromosomes parents dans le but d'enrichir la population. Dans notre travail, nous choisissons aléatoirement deux points de croisement pour effectuer la concaténation de deux fragments du premier chromosome avec un fragment du deuxième chromosome. Ainsi un nouveau chromosome résultant de cette concaténation est inséré dans la population en remplacement d'un chromosome n'ayant pas un bon fitness.

4. Description de notre algorithme

Donnees :

N: nombre de sous-porteuses
 K: nombre d'utilisateurs
 T: taille de la population
 Ngen: nombre de generation

```
Initialisation_population(population,N,K)
Evaluation_chromosome(population,N,K)
for i=1 to Ngen
{
  Selection_tournament(population,N,K)
  Mutation(population,N,K)
  Crossover(population,N,K)
}
search_best_chromosome(population)
ajuste_allocation(best_chromosome)
```

Contrairement à certaines implémentations dans lesquelles la notion de chromosomes valides et invalides est définie ; ce qui peut affecter la diversité de la population. La particularité de notre implémentation est la fonction *ajuste_allocation()*. Cette dernière permet d'ajuster le schéma d'allocation du meilleur chromosome résultant de l'algorithme génétique en retirant le surplus de sous-porteuses aux utilisateurs ayant un débit supérieur à la contrainte et en le réaffectant aux utilisateurs avec un déficit de débit.

5. Simulation

5.1. Paramètres simulation

- BER(Bit Error Rate) : 10^{-2} to 10^{-4}
- M-QAM : $M \in \{4, 8, 16, 32, 64\}$
- $N_0 = 0.01$: Densité spectrale de puissance du bruit
- Nombre de sous-porteuses : 128
- Taille Population : 200
- Rapidité modulation : 14kbauds
- Débit Minimal par utilisateur : 750kbits/s

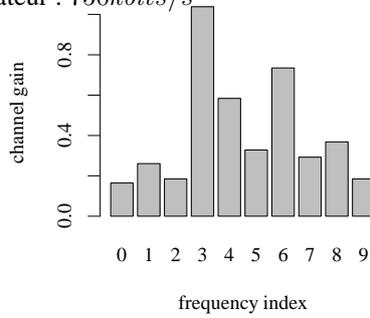
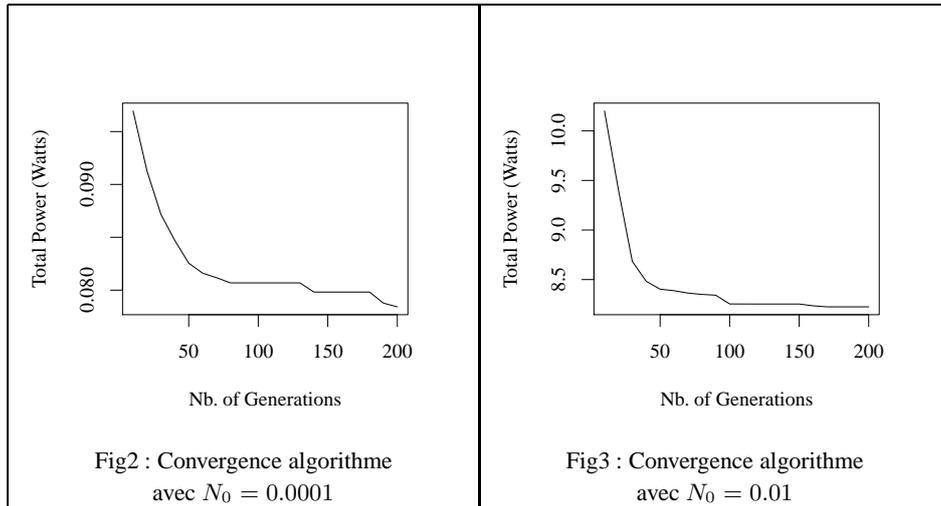


Fig1 : Variation du gain canal d'un utilisateur.

5.2. Résultats simulation



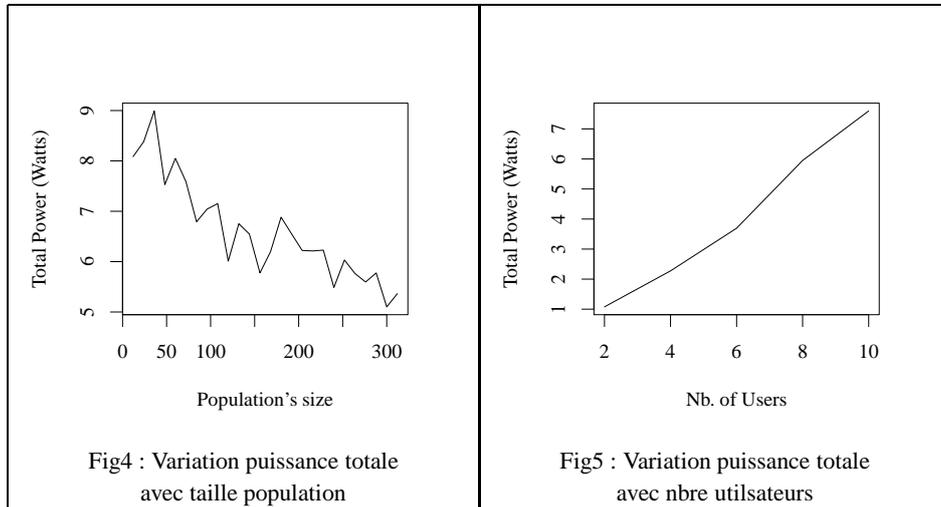


Fig4 : Variation puissance totale avec taille population

Fig5 : Variation puissance totale avec nbre utilisateurs

Tableau 1 : Comparaison de notre algorithme avec l'algorithme proposé par [1] pour $N_0 = 0.0001$

	2 util.	4 util.	8 util.
Notre Algorithme	0.01w	0.022w	0.061w
Algorithme [1]	0.03 w	0.05 w	0.12 w

Tableau 2 : Comparaison de notre algorithme avec l'algorithme proposé par [10] pour $N_0 = 0.01$

	2 util.	4 util.	6 util.	8 util.
Notre Algorithme	0.98 w	2.09 w	3.77 w	5.68 w
Algorithme[10]	5w	8 w	12w	15w

5.3. Analyse des résultats

L'étude de l'évolution de la puissance totale en fonction du nombre de génération (figure 2 et 3) montre que la puissance reste relativement constante à partir de la 80e génération et le schéma d'allocation correspondant est la solution au problème d'allocation de ressources. La figure 4 montre quelques problèmes de convergence de notre algorithme ; cependant la puissance totale continue de diminuer pour une taille croissante de la population. Notons que notre implémentation donne de meilleurs résultats par rapport aux principaux algorithmes vus dans la littérature. Les tableaux 1 et 2 ci-dessus illustrent cette comparaison.

6. Conclusion

Dans cet article nous avons proposé une implémentation de l’algorithme génétique pour la résolution du problème d’allocation de ressources dans les systèmes multi-utilisateurs OFDM. Notre implémentation est caractérisée par une population et des chromosomes de taille fixe. De plus nous avons défini une fonction supplémentaire *ajuste_allocation()* permettant d’équilibrer le schéma d’allocation. Après simulation, notre algorithme a offert de meilleurs résultats que les algorithmes proposés par [1] et [10]. Dans nos futurs travaux nous envisageons d’apporter des améliorations sur certains opérateurs de l’algorithme génétique en vue d’améliorer la convergence de notre implémentation. Nous comptons aussi faire l’étude comparative de la complexité de notre algorithme avec celle des algorithmes de référence.

7. Bibliographie

- [1] H. Ahmadi and Y. Chew. Adaptive subcarrier-and-bit allocation in multiclass multiuser ofdm systems using genetic algorithm. IEEE, 2009.
- [2] M. Alouini and A. Goldsmith. Adaptive m-qam modulation over nakagami fading channels. *IEEE, Global Communication Conf.*, 1997.
- [3] S. Plefischinger G. Münz and J. Speidel. An efficient waterfilling algorithm for multiple access ofdm. Stuttgart Germany, 2002. IEEE Globecom 2002.
- [4] David Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. MA Addison Wesley, USA, 1th edition, 1989.
- [5] GRETSI. Rapport colloque gretsi sur l’ofdm. Technical report, Laboratoire ETIS CNRS, Université de Cergy-Pontoise, France, 2007.
- [6] T. Ibaraki and N. Katoh. Resource allocation problems-algorithmic approaches. Technical report, MIT Press, 1988.
- [7] Mélanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, USA, 1th edition, 1999.
- [8] John Proaski. *Digital Communications*. McGraw Hill, New York, USA, 4th edition, 2001.
- [9] Qilin Qi and Y.L.Yang. Efficient water-filling algorithm for power allocation in ofdm-based cognitive radio systems. *CSE Conference and Workshop Papers*, (196), Janvier 2012.
- [10] Y. Reddy and N.Gajendar. Evolutionary approach for efficient resource allocation in multi-user ofdm systems. Las Vegas Nevada USA, 2007. Journal of Communications IEEE.
- [11] H. Rohling. *Ofdm : Concepts for future communication systems*. Hamburg Germany, 2011. Springer.
- [12] H. Schulze and C. Lüders. *Theory and Application of OFDM and CDMA*. John Wiley & Sons Ltd, Germany, 1th edition, 2005.
- [13] Wong and al. al. Multiuser ofdm with adaptive subcarrier, bit, and power allocation. *IEEE*, 17(10), Octobre 1999.

=====

Réutilisation de ressources terminologiques existantes pour la construction d'une ontologie de domaine

Khadim Dramé, Gayo Diallo, Fleur Mougin

LESIM-ISPED
Université de Bordeaux
33000 Bordeaux
FRANCE

Prenom.Nom@isped.u-bordeaux2.fr

=====

RÉSUMÉ. Nous présentons dans ce papier une méthode de construction d'ontologies (bilingues) à partir de ressources existantes du domaine. Cette méthode combine deux approches : l'acquisition de connaissances à partir de textes et la réutilisation de ressources terminologiques existantes du domaine. L'approche comprend quatre étapes : i) l'extraction des candidats termes dans les corpus français et anglais en utilisant des outils d'analyse de textes, ii) le regroupement des termes en concepts et leur structuration en se basant sur UMLS, iii) l'appariement des termes français et anglais, iv) le raffinement et la validation des résultats par des spécialistes du domaine. Après validation, les concepts sont formalisés au sein d'une ontologie de domaine.

ABSTRACT. In this paper, we present a method for building (bilingual) domain ontologies from existing resources. This method combines two approaches: knowledge extraction from texts and the reuse of existing terminological resources. The approach consists of four steps: i) the extraction of terms from French and English corpus using textual analysis tools, ii) terms clustering in concepts and concepts structuring based on the UMLS, iii) the alignment of French and English terms, iv) the refinement and validation of results by domain experts. These validated concepts are formalized so that into a domain ontology.

MOTS-CLÉS : construction d'ontologies à partir de textes, UMLS, maladie d'Alzheimer, réutilisation de ressources existantes.

KEYWORDS: ontologies learning from text, UMLS, Alzheimer's disease, existing resources reuse.

=====

*Actes du 5^e Colloque National sur la Recherche en Informatique et ses Applications – Ziguinchor,
Sénégal, Avril 2013*

=====

1. Introduction

Le projet SemBiP (Semantic BiblioDem Portal) vise à fournir une stratégie pour la structuration et l'exploitation efficace des connaissances sur la maladie d'Alzheimer et les syndromes apparentés. Le premier objectif du projet consiste à modéliser les connaissances du domaine sous forme d'une ontologie. Cette ontologie se veut bilingue pour pouvoir supporter un portail de recherche d'information sémantique inter-lingue. Même si, dans le domaine biomédical, d'importantes ressources termino-ontologiques ont été développées, il y a encore des besoins pour couvrir de manière plus exhaustive des domaines spécifiques comme la maladie d'Alzheimer dont les connaissances évoluent rapidement et sont régulièrement enrichies. Par ailleurs, les larges ressources telles qu'UMLS manquent de sémantique formelle et sont principalement en anglais, ce qui handicape leur utilisation dans un contexte francophone notamment. La modélisation des connaissances du domaine de la maladie d'Alzheimer reste aujourd'hui un enjeu important pour une meilleure connaissance de celle-ci. Les travaux en ingénierie des connaissances peuvent servir à une telle modélisation autour de la notion d'ontologie. Une ontologie permet de représenter explicitement les connaissances d'un domaine au moyen de concepts et des relations qu'ils entretiennent [1] et facilite ainsi leur partage et leur exploitation de manière automatique. La construction d'ontologies reste toutefois une tâche délicate et chronophage. Même si plusieurs approches méthodologiques ont été proposées pour guider ce processus [2], la plupart d'entre elles sont manuelles; ce qui rend leur application difficile et chronophage. D'autres approches proposent d'alléger le processus en automatisant certaines étapes. Parmi ces dernières, les méthodes d'acquisition d'ontologie à partir de textes [3] ont été largement utilisées ces dernières années dans l'ingénierie ontologique. Ces approches considèrent les textes comme principales sources de connaissances et s'appuient sur des méthodes de traitement automatique des langues (TAL). Enfin, compte tenu des efforts requis pour la construction d'ontologies, la réutilisation de ressources existantes du domaine reste une question très importante.

Ainsi, dans le cadre de notre travail, nous avons opté pour une approche qui se base sur des corpus représentatifs du domaine extraits à partir de la base BiblioDem¹ et la réutilisation d'une large ressource du domaine (UMLS) pour assister la tâche de structuration des connaissances. Concrètement, notre méthode utilise des outils de TAL pour repérer les termes candidats puis repose sur UMLS pour regrouper ces termes en concepts et les organiser via des relations sémantiques. Cette ontologie est ensuite enrichie par l'ajout de nouveaux concepts et relations pertinents du domaine en utilisant des relations syntaxiques (cf. section 3.4).

¹ http://www.isped.u-bordeaux2.fr/CDD/FR_HTML_BIBLIONET.aspx#BiblioDem

2. Ressources

2.1. UMLS

UMLS[®] (Unified Medical Language System[®]) est composé principalement de deux ressources sémantiques : le Metathesaurus[®] et le réseau sémantique [4]. Le Metathesaurus de UMLS est un large graphe constitué de plus de deux millions de concepts, définis par intégration d'une centaine de terminologies biomédicales. Chaque concept du Metathesaurus est un cluster de termes synonymes, éventuellement définis dans des langues différentes, ayant un identifiant unique (Concept Unique Identifier – CUI) et auquel est parfois associée une définition. Ces concepts sont reliés par des relations dont la plupart sont issues des terminologies sources. Ces relations sont hiérarchiques ou transversales et certaines d'entre elles sont « définies » d'un point de vue logique (c'est-à-dire qu'elles sont typées comme par exemple *is_a*, *finding_site_of*, *part_of*). Le réseau sémantique est composé d'une hiérarchie de 133 types sémantiques. Chaque concept du Metathesaurus est catégorisé par un ou plusieurs types sémantiques (par exemple, *Alzheimer'sdisease* est catégorisé par le type sémantique *Disease or Syndrome*).

2.2. Corpus de travail

BiblioDem est une base cumulative qui contient environ 1 300 articles scientifiques concernant la maladie d'Alzheimer et les syndromes apparentés. Cette base consiste en l'analyse critique d'articles scientifiques de la littérature mondiale sur la maladie et ses syndromes apparentés. Elle est enrichie chaque mois d'articles sélectionnés dans des bases documentaires comme MEDLINE et de leurs analyses critiques. Nous avons constitué deux corpus textuels à partir de cette base représentative du domaine : un corpus anglais (307 597 mots) composé des abstracts des articles et un corpus français (641 612 mots) regroupant les synthèses et commentaires de ces mêmes articles.

3. Méthode

Notre méthodologie comprend quatre étapes. La première étape consiste en l'acquisition automatique de termes à partir de ces corpus grâce à l'outil d'extraction terminologique Syntex et le filtrage de ses résultats. La deuxième étape se focalise sur la conceptualisation de ces termes candidats à l'aide d'une ressource externe : UMLS. Pour enrichir les connaissances disponibles en langue française, nous proposons ensuite une méthode d'appariement de termes. La dernière étape concerne la validation des résultats par des spécialistes du domaine.

3.1. Extraction de termes

Pour extraire les candidats termes à partir de ces corpus, nous avons utilisé l'outil Syntex. Le choix de cet outil est motivé par sa performance et sa capacité à traiter à la fois des textes anglais et français. Syntex est un analyseur syntaxique qui permet d'extraire des candidats termes (noms, syntagmes nominaux) et des relations syntaxiques les structurant à partir de corpus de textes [5]. A partir d'un texte étiqueté préalablement par un outil tel que TreeTagger², il fournit un réseau de dépendances syntaxiques constitué de mots et syntagmes avec leur contexte d'apparition. Par exemple le terme *déclin cognitif léger* a pour tête *déclin cognitif*.

L'extracteur fournit des listes de candidats termes (CT) bruitées nécessitant un filtrage automatique pour obtenir un résultat plus pertinent. Ainsi, dans un premier temps, nous considérons seulement les CT les plus représentatifs, i.e., ayant une fréquence d'apparition dans le corpus supérieure à un seuil (ici, nombre d'occurrences ≥ 7). Ensuite, les mots vides³ et les termes constitués exclusivement de chiffres sont élagués, les CT de longueur inférieure à quatre et les CT contenant des chiffres ou des caractères non alphanumériques sont isolés pour vérification.

3.2. Conceptualisation

Cette étape consiste à regrouper les termes synonymes en concepts et à les organiser via des relations. La conceptualisation manuelle demeurant une tâche coûteuse et fastidieuse, nous avons utilisé UMLS pour automatiser cette tâche. Nous avons ainsi utilisé le service Web d'UMLS qui permet de trouver le(s) concept(s) du Metathesaurus correspondant à un terme. Nous avons tout d'abord effectué une recherche exacte sur la liste de tous les termes retenus et pour ceux qui n'ont pas été retrouvés tels quels lors de cette première étape, une recherche normalisée a été réalisée. La méthode de recherche normalisée prend en compte les variations lexicales des termes (flexion, dérivation) en se basant sur le programme Lexical Variant Generation⁴. La chaîne d'entrée est d'abord normalisée en un ensemble de variantes qui sont ensuite comparées aux termes du Metathesaurus. Ainsi, la recherche normalisée permet d'associer les termes *Parkinson's disease dementia* et *Older adults* respectivement aux concepts *Dementia in Parkinson's disease* (C0349081), *Older Adulthood* (C1999167).

En plus des concepts, les relations de subsumption (typées *is_a*), directes comme indirectes via des concepts intermédiaires n'appartenant pas à la liste des concepts trouvés, sont récupérées. L'ensemble de ces concepts intermédiaires ainsi que tous les concepts liés aux concepts spécifiques du domaine par des relations taxonomiques

² <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

³ Listes française et anglaise disponibles sur <http://nlp.cs.nyu.edu/GMA/docs/resources.html>

⁴ http://www.nlm.nih.gov/research/umls/new_users/online_learning/LEX_004.htm

explicités sont aussi intégrés pour enrichir et mieux structurer l'ontologie. Pour les concepts n'entretenant pas de relations taxonomiques avec d'autres concepts, les relations plus générales (*Child*, *Narrower*) sont considérées comme des relations hiérarchiques candidates, à valider dans un deuxième temps. Différents types de relations transversales exprimées de manière logique entre les concepts sont également extraites (par exemple, *treat*, *cause_of*, *anatomical_part_of*).

Nous avons développé un algorithme de suppression automatique des relations hiérarchiques redondantes (voir figure 1). En plus de ces redondances, des cycles ont été étudiés manuellement. Pour les couples de concepts reliés par différents types de relations, nous avons préservé uniquement la relation sémantique la plus spécifique (hiérarchie des relations établie manuellement).

Enfin, pour chaque concept obtenu, nous avons extrait d'UMLS ses éventuels synonymes et définitions en anglais et en français. Le portail terminologique de santé de CISMef⁵ intégrant plusieurs ressources terminologiques en français, il a été utilisé pour enrichir l'ontologie avec des termes synonymes français. Pour organiser les concepts à un plus haut niveau, les types sémantiques les catégorisant ont également été intégrés dans l'ontologie.

Chaque concept est ainsi représenté par une classe ayant pour propriétés un URI, un terme préféré en anglais et parfois en français, et éventuellement des termes synonymes et une définition, et enfin le(s) type(s) sémantique(s) le catégorisant.

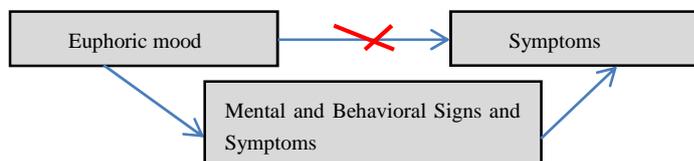


Figure 1 : Exemple de relation hiérarchique redondante

3.3. Appariement des termes

La recherche des concepts associés aux termes dans UMLS met en évidence une large prédominance de l'anglais dans cette ressource car les terminologies qui y sont intégrées sont principalement en anglais. Pour compléter la partie française et développer ainsi une ressource bilingue riche du domaine, nous avons utilisé une technique d'alignement des termes. Cette méthode d'alignement s'inscrit dans le cadre des approches d'alignement à partir de corpus parallèles qui sont largement utilisées pour la construction de ressources terminologiques multilingues [6]. Elles consistent en la mise en correspondance de termes extraits dans des textes parallèles.

⁵ <http://www.chu-rouen.fr/cismef/>

Dramé et al.

Un corpus parallèle a été constitué à partir des titres des articles en anglais et leur traduction en français. Pour aligner les termes repérés dans le texte source avec leurs équivalents dans le texte traduit, la méthode proposée combine deux techniques [7] : une technique heuristique basée sur le calcul de score d'association et une méthode linguistique fondée sur la similarité morphologique. Le score d'association choisi, l'indice de Jaccard, est couramment utilisée dans la littérature ; en plus d'être simple, il permet de déterminer les couples de termes les plus fortement associés dans un corpus aligné. Pour vérifier la réciprocity de l'équivalence entre termes, nous avons réalisé l'appariement dans les deux sens en considérant d'une part l'anglais comme langue source et le français comme langue cible et inversement. Les termes médicaux (pour les langues française et anglaise) possédant de nombreuses racines gréco-latines, notre méthode linguistique s'appuie sur la distance de Levenshtein normalisée qui permet de repérer les couples de termes identiques ou similaires dans les deux langues.

3.4. Validation

L'ensemble des concepts trouvés dans UMLS ont été présentés aux experts du domaine (deux spécialistes de la maladie d'Alzheimer) avec éventuellement les termes associés et leur contexte d'apparition dans les corpus afin qu'ils jugent de leur pertinence. Les relations taxonomiques candidates ont été aussi analysées pour compléter la hiérarchie des concepts. Les résultats de l'appariement et les termes non trouvés dans UMLS ont également été présentés aux experts (une interne en santé publique et une doctorante de l'équipe Epidémiologie et Neuropsychologie du Vieillissement Cérébral). Ceux qu'ils ont jugés pertinents ont été par la suite intégrés dans l'ontologie grâce aux dépendances syntaxiques fournies par Syntex.

4. Résultats

Syntex a fourni 49 390 syntagmes nominaux (SN) et 8 844 termes simples pour le corpus anglais et 69 505 SN et 11 688 termes simples pour le corpus français. Après filtrage, nous avons obtenu des listes largement réduites de CT (6%). Des exemples de termes supprimés sont *effet*, *years*, *information*, *importance*. En revanche, des termes tels que *Vitamine B12*, *Age*, *DNA* ayant été initialement isolés ont finalement été ajoutés aux CT car ils ont été jugés pertinents par les experts du domaine.

La recherche des termes dans UMLS montre qu'une bonne partie des termes simples (65%) retenus est contenue dans cette ressource alors que moins d'un tiers des SN (27%) l'est. Ce résultat s'explique par le fait que les SN sont des termes spécifiques au domaine et sont donc moins présents dans UMLS. Des exemples de termes non trouvés dans UMLS sont *cognitive tests*, *neuropsychiatric symptoms*. Enfin, la prédominance des termes anglais dans UMLS est clairement mise en évidence (45% contre 34%).

CNRIA'2013

La fusion de l'ensemble de ces résultats a permis d'identifier 3 871 concepts distincts dans UMLS. Après validation par les experts, 2 421 concepts ont été retenus. Par exemple, les concepts *Alzheimer's Disease* (C0002395), *Dementia* (C0497327), *risk factors* ont été jugés valides. Par contre, les concepts *Scientific Control* (C1882979), *Science of Anatomy* (C0002808) et *Number of patients* (C2360800) n'ont pas été gardés. En plus de ces concepts, 3 028 concepts supplémentaires ont été intégrés afin de structurer les concepts au sein de l'ontologie. Par exemple, via les concepts *Alzheimer's disease* et *nervous system disorder* (C0027765) initialement trouvés, le concept intermédiaire *Cerebral degenerations* (C2026458) a également été ajouté.

L'ontologie a été enrichie par la suite par l'intégration de 436 nouveaux concepts et 908 synonymes français supplémentaires grâce à l'exploitation du corpus parallèle de titres. Des exemples de synonymes obtenus par appariement sont *Parkinson's disease* (*maladie de Parkinson*), *family history of dementia* (*antécédents familiaux de démence*). Un exemple de nouveau concept est *severe Alzheimer's disease* que l'on a intégré comme fils de *Alzheimer's disease* grâce aux dépendances fournies par Syntex.

5. Discussion

La réutilisation de ressources existantes est une question importante notamment dans le domaine médical où de larges ressources terminologiques sont disponibles. L'approche proposée repose sur UMLS pour alléger le processus de construction d'ontologie. Cependant, malgré sa richesse, UMLS est limitée par son niveau de spécification superficiel avec des relations généralement vagues (*Broader*, *Related*, etc.). De plus, la variété de ses sources et sa large couverture conduisent à des redondances, des cycles et parfois même à des incohérences. Ainsi, des mécanismes ont été développés pour l'élagage des redondances, la correction des incohérences et la spécification explicite des relations. Ensuite, une phase d'enrichissement de l'ontologie avec l'ajout de nouveaux concepts et synonymes est réalisée. L'ensemble de ces résultats a été raffiné et validé par des spécialistes du domaine. Dans la littérature, beaucoup de travaux se sont intéressés à la réutilisation de ressources existantes dans l'ingénierie ontologique. Une méthodologie générale proposée dans [8] définit une technique de réutilisation des ressources ontologiques et non-ontologiques pour alléger le processus de développement d'ontologies. Elle propose une procédure de transformation de ces ressources de connaissances en ontologies en utilisant WordNet pour rendre explicite les relations. Dans le domaine médical, Hahn et Schulz [9] ont également proposé une méthodologie similaire pour construire une ontologie formelle à partir de UMLS. Les auteurs se sont basés sur les connaissances contenues dans cette ressource pour concevoir un modèle formel avec un niveau de conceptualisation plus détaillé. Mais leur approche nécessite des efforts supplémentaires de modélisation et un coût de validation manuelle. Chrisment et ses collègues [10] s'appuient sur un thésaurus

et un corpus de référence pour construire une ontologie dans le domaine de l'astronomie. Pour structurer les concepts, ils considèrent les relations *<est plus spécifique>* et *<est plus générique>* qui sont plus vagues que la relation de subsomption. En plus, ils ne proposent pas de mécanismes automatiques pour désambiguïser ces relations. Ce travail est proche des travaux de [10] mais, pour la hiérarchie des concepts, nous considérons seulement les relations taxonomiques explicites. En outre, puisque nous disposons d'une ressource suffisamment fournie (UMLS), nous avons exploité en plus du réseau sémantique (extraction de types sémantiques), des relations associatives spécifiques du Metathesaurus pour mieux structurer les concepts. Par rapport aux travaux de Hahn et ses collègues, au-delà des relations taxonomiques et partitives, nous exploitons d'autres relations sémantiques d'intérêt. En plus, notre approche comprend une phase d'enrichissement avec l'intégration de nouvelles entités sémantiques.

6. Bibliographie

- [1] T. Gruber, « A translation approach to portable ontology specifications », *Knowledge Acquisition*, vol. 5, p. 199-220, juin 1993.
- [2] F. Lopez, « Overview Of Methodologies For Building Ontologies », 1999.
- [3] N. Aussenac-Gilles, S. Despres, et S. Szulman, « The TERMINAE Method and Platform for Ontology Engineering from Texts », in *Proc. of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, 2008, p. 199-223.
- [4] O. Bodenreider, « The Unified Medical Language System (UMLS): integrating biomedical terminology », *Nucleic Acids Research*, vol. 32, p. 267-270, 2004.
- [5] D. Bourigault et C. Fabre, « Approche linguistique pour l'analyse syntaxique de corpus », *Cahiers de Grammaire - Université Toulouse le Mirail*, n° 25, p. 131-151, 2000.
- [6] F. J. Och et H. Ney, « A systematic comparison of various statistical alignment models », *Comput. Linguist.*, vol. 29, n° 1, p. 19-51, mars 2003.
- [7] K. Dramé, G. Diallo, et F. Mougin, « Towards a bilingual Alzheimer's disease terminology acquisition », *Stud Health Technol Inform.* 2012; 180:179-183.
- [8] A. García-Silva, A. Gómez-Pérez, M. C. Suárez-Figueroa, et B. Villazón-Terrazas, « A Pattern Based Approach for Re-engineering Non-Ontological Resources into Ontologies », in *Proc. of the 3rd Asian Semantic Web Conference on The Semantic Web*, Berlin, Heidelberg, 2008, p. 167-181.
- [9] U. Hahn et S. Schulz, « Building a very large ontology from medical thesauri », *Handbook on Ontologies*, p. 133-150, 2004.
- [10] C. Chrisment, O. Haemmerlé, N. Hernandez, et J. Mothe, « Méthodologie de transformation d'un thesaurus en une ontologie de domaine », *Revue d'Intelligence Artificielle*, vol. 22, n° 1, p. 7-37, 2008.

.....

Amélioration de la consommation d'énergie fondée sur la redondance de mesure dans les WSNs

Ibrahima Diané¹, Rahim Kacimi¹, Zoubir Mammeri¹, Ibrahima Niang²

¹ IRIT-UPS, Université de Toulouse, France
{diane, kacimi, mammeri}@irit.fr

² LID-UCAD, Université de Dakar, Sénégal
iniang@ucad.sn

.....

RÉSUMÉ. Cet article met l'accent sur la prise en charge de la redondance de mesure dans les réseaux de capteurs organisés en clusters. La plupart des travaux sur ces réseaux s'intéressent à la consommation d'énergie et ne prennent pas en compte la notion de redondance de mesure. Cette redondance est aussi une source inutile de consommation d'énergie. Cet article propose MTMR (*Management Technique of the Measurement Redundancy*), une technique de regroupement de nœuds capteurs dans les clusters. Il introduit la notion de proximité géographique pour prendre en compte la redondance. Cette technique améliore considérablement la consommation d'énergie et garantit un meilleur délai de bout en bout.

Les résultats de simulation montrent que notre approche permet de réduire considérablement la consommation d'énergie dans les réseaux de capteurs et augmente ainsi leur durée de vie. Il diminue aussi le délai de bout en bout.

ABSTRACT. This paper focuses on measurement redundancy in wireless sensor networks organized in clusters. Most of the work on WSN does not take into account the measurement redundancy. This redundancy is also an unnecessary source of energy consumption. We propose MTMR (*Management Technique of the Measurement Redundancy*), a grouping technique for sensor nodes in the clusters. It introduces the notion of *geographical proximity* to take into account the redundancy. This technique significantly improves the energy consumption and ensures a better end to end delay.

The simulation results show that our approach reduces greatly the energy consumption in WSNs and thus increases their lifetime. It also improves the end to end delay.

MOTS-CLÉS : réseaux de capteurs, hiérarchisation, redondance de mesure, proximité géographique.

KEYWORDS: WSN, clustering, measurement redundancy, geographical proximity.

.....

1. Introduction

Un réseau de capteurs se définit comme un ensemble de capteurs connectés entre eux, chaque capteur étant muni d'un émetteur-récepteur. Les réseaux de capteurs forment une nouvelle génération de réseaux aux propriétés spécifiques, qui n'entrent pas dans le cadre des architectures classiques. La miniaturisation des capteurs pose des problèmes de communication et de ressources d'énergie. Il faut que le capteur soit suffisamment intelligent pour collecter ou recevoir l'information et l'émettre à faible consommation d'énergie avec un délai raisonnable. Le déploiement d'un réseau de capteurs pose donc des problèmes de gestion de l'alimentation et de délai de bout en bout. Malgré l'introduction de la hiérarchisation, les réseaux de capteurs sont toujours confrontés à des problèmes de consommation d'énergie et de délai. La plupart des travaux [2, 3, 5, 6, 8, 9, 11, 12, 13, 16] propose des améliorations en ce qui concerne la consommation d'énergie. Cependant, ils ne prennent pas en compte la redondance de mesure. Ces travaux utilisent généralement l'agrégation de données et augmentent le délai.

Lorsque la taille du réseau devient importante, sa gestion soulève des difficultés. La structuration d'un réseau de capteurs est un des outils principaux pour sauvegarder l'énergie dans chaque nœud du réseau, ce qui permet de prolonger la vie du système. Une des structures les plus connues est la hiérarchie. La technique de hiérarchisation sert à partitionner le réseau en sous ensembles afin de faciliter sa gestion. Dans ce type de techniques, la vue du réseau devient locale; des nœuds spéciaux peuvent avoir des rôles supplémentaires. La littérature comprend plusieurs contributions sur les techniques de hiérarchisation du réseau, que nous allons présenter brièvement. Nous distinguons deux types de groupes de nœuds : la zone et le cluster. Un cluster est défini par un ensemble de nœuds et possède un nœud nommé nœud-chef ou Cluster Head (CH). Le rôle du CH est de faire le relais entre les nœuds du cluster et la station de base directement ou via d'autres CHs. Le CH possède généralement des ressources énergétiques supérieures aux autres nœuds du réseau. Le CH est élu suivant différents critères et informations sur le réseau : le niveau de l'énergie, la connexion avec les autres capteurs, la position géographique, etc. Une zone est définie par un ensemble de nœuds mais ne possède pas un CH. Ainsi, un cluster est une sous-classe d'une zone. La formation des groupes (zones ou clusters) s'appuie sur des informations sur le réseau, exigeant donc son instrumentation.

Dans cet article, nous proposons une technique de regroupement de capteurs redondants dans les réseaux de capteurs hiérarchiques. Cette technique prend en compte la redondance de mesure dans les réseaux de capteurs. Notre approche s'appuie sur un schéma hiérarchique à trois niveaux comme dans LEACH [6].

Le reste de cet article est organisé comme suit. La section 2 présente les techniques de hiérarchisation dans les réseaux de capteurs. La section 3 présente notre approche de regroupement par proximité géographique. Une validation expérimentale est proposée dans la section 4. Enfin, une conclusion est formulée dans la section 5.

2. État de l'art

Dans [6], Heinzelman et al. ont proposé un algorithme de hiérarchisation distribué appelé LEACH pour les réseaux de capteurs homogènes. LEACH choisit aléatoirement les nœuds CHs et attribue ce rôle aux différents nœuds selon la politique de gestion Round-Robin pour garantir une dissipation équitable d'énergie entre les nœuds. Dans le but de réduire la quantité d'informations transmises à la station de base, les CHs agrègent les données capturées par les nœuds membres qui appartiennent à leur propre cluster, et envoient un paquet agrégé à la station de base. LEACH est exécuté en deux phases : la phase *set-up* et la phase *steady-state*. Dans la première phase, les CHs sont sélectionnés et les clusters sont formés, et dans la seconde phase, le transfert de données vers la station de base aura lieu. Cependant, bien que LEACH puisse augmenter la durée de vie du réseau, il présente certaines limitations. LEACH suppose que tous les nœuds puissent transmettre des données avec une grande puissance pour atteindre la station de base et que chaque nœud a une puissance de calcul lui permettant de supporter différentes

couches MAC. Par conséquent, LEACH ne convient pas aux réseaux déployés dans de vastes régions. En outre, dans LEACH, l'agrégation des données est centralisée et exécutée périodiquement par le CH. Cette agrégation augmente également le délai de bout en bout.

Une variante de LEACH appelée LEACH-C [4, 5] a été conçue pour améliorer les performances de LEACH. Cette variante utilise une architecture centralisée pour choisir les CHs tout en impliquant la station de base et l'information de localisation des capteurs. Cependant, elle augmente considérablement le surcoût du réseau puisque tous les capteurs devront envoyer leurs informations de localisation à la station de base en même temps pendant chaque phase d'élection de CHs. D'autre part, le délai n'est toujours pas amélioré. A l'instar de [14], plusieurs travaux ont prouvé qu'une telle architecture centralisée ne supporte pas le passage à l'échelle, étant plus particulièrement appropriée à des réseaux de petite taille.

Dans [11], Lindsey et Raghavendra ont proposé une version améliorée de LEACH appelée PEGASIS. L'idée principale de PEGASIS est de former une chaîne entre les nœuds de sorte que chaque nœud reçoive de (et communique à) un voisin proche. Les données collectées sont transmises d'un nœud à un autre qui les agrège jusqu'à ce qu'elles arrivent à un nœud particulier qui les transmet à la station de base. Les nœuds qui transmettent les données à la station de base, sont choisis tour à tour selon une politique round-robin dans le but de réduire l'énergie moyenne dépensée par un nœud durant une période (round). Cependant, l'agrégation des données au niveau de tous les nœuds relais, cause une augmentation considérable du délai de bout en bout.

Manjeshwar et Agrawal [12] ont proposé une technique de clustering appelée TEEN pour les applications critiques où le changement de certains paramètres peut être brusque. L'architecture du réseau est basée sur un groupement hiérarchique à plusieurs niveaux où les nœuds les plus proches forment des clusters. Puis ce processus de clustering passe au deuxième niveau jusqu'à ce que la station de base soit atteinte. Après la formation des clusters, chaque CH transmet à ses membres deux seuils : un seuil Hard HT (Hard Threshold), qui est la valeur seuil du paramètre contrôlé (surveillé) et un seuil Soft ST (Soft Threshold) représentant une petite variation de la valeur du paramètre contrôlé. L'occurrence de cette petite variation ST permet au nœud qui la détecte de la signaler à la station de base en transmettant un message d'alerte. Par conséquent, le seuil Soft réduira le nombre de transmissions puisqu'il ne permet pas la transmission s'il y a peu ou pas de variation de la valeur du paramètre contrôlé. Puisque la transmission d'un message consomme plus d'énergie que la détection des données, alors la consommation d'énergie dans TEEN est moins importante que dans les protocoles qui transmettent des données périodiquement tels que LEACH. Cependant, l'inconvénient principal de ce protocole est que, si les seuils HT et ST ne sont pas reçus, les nœuds ne communiqueront jamais, et aucune donnée ne sera transmise à l'utilisateur, ainsi la station de base ne connaît pas les nœuds qui ont épuisé leur énergie. TEEN ne convient pas aux applications qui nécessitent des envois périodiques de données.

Pour remédier aux limitations du protocole TEEN, les auteurs ont proposé une extension de TEEN appelée APTEEN [13]. APTEEN est un protocole hybride qui change la périodicité et les valeurs seuils utilisées dans TEEN selon les besoins de l'utilisateur et le type d'application. Dans APTEEN, les CHs transmettent à leurs membres les paramètres suivants : l'ensemble de paramètres physiques auxquels l'utilisateur est intéressé pour obtenir des informations (A) ; les seuils : seuil Hard HT et seuil Soft ST ; un Schedule TDMA permettant d'assigner à chaque nœud un intervalle fini de temps appelé slot ; un Compteur de Temps (CT) : c'est la période de temps maximum entre deux transmissions successives d'un nœud. APTEEN offre une grande flexibilité qui permet à l'utilisateur de choisir l'intervalle de temps CT, et les valeurs seuils HT et ST pour que la consommation d'énergie soit contrôlée par la variation de ces paramètres. Cependant, APTEEN nécessite une complexité supplémentaire pour implémenter les fonctions de seuils et de périodes de temps CT. Ainsi, le surcoût et la complexité associés à la formation des clusters à plusieurs niveaux par TEEN et APTEEN sont assez élevés.

3. Redondance dans les réseaux de capteurs sans fil

3.1. Problématique

La plupart des travaux proposent des améliorations en ce qui concerne la consommation d'énergie sans tenir compte la redondance de mesure qui est également une source de consommation inutile d'énergie. La description du placement initial peut expliciter une certaine redondance dans les capteurs déployés. Les réseaux de capteurs sont caractérisés par un nombre important de nœuds. La redondance de mesure peut découler de cette forte densité de nœuds. Plusieurs nœuds peuvent détecter un même événement et donc reporter la même valeur de mesure. Il faut donc prévoir des mécanismes pour la prise en charge d'éventuelles redondances de mesure. La problématique est donc ici la prise en compte de la redondance de mesure dans les réseaux de capteurs pour une meilleure amélioration de la consommation d'énergie.

Cet article propose et analyse une technique pour la prise en compte de la redondance de mesure dans les réseaux de capteurs hiérarchiques.

3.2. Technique de regroupement par proximité géographique

Les architectures en clusters sont largement utilisées dans les WSNs pour leur capacité de passage à l'échelle. Cette structuration doit permettre une utilisation plus efficace des ressources rares du réseau telles que l'énergie et la bande passante.

MTMR est utile uniquement dans le cas de mesures périodiques. Cette technique permet de regrouper les nœuds feuilles qui, en fonction de leur position géographique, captent la même valeur de mesure. Ainsi, les communications intra-clusters sont réduites et la consommation d'énergie est optimisée.

3.2.1. Hypothèses

Pour faciliter la compréhension de notre approche dans le reste de ce document, nous faisons quelques hypothèses raisonnables :

- Comme LEACH, une architecture réseau à trois niveaux.
- Les nœuds sont aléatoirement et densément déployés.
- Les nœuds observent le même phénomène physique et génèrent un trafic périodique.
- Aucune information sur une éventuelle redondance n'est explicitement fournie.

3.2.2. Principe de notre approche

MTMR combine la proximité géographique des nœuds et la redondance de mesure pour réduire le nombre de transmissions intra-clusters. L'initialisation du protocole MTMR comprend deux principales étapes.

- Dans la première étape, les CHs collectent et observent les données dans leur cluster. Après le déploiement des nœuds et la formation des clusters, chaque CH collecte les données des nœuds. Cette phase est dimensionnée par la durée *IntMeasure*.
- Dans la deuxième étape, les CHs déterminent la redondance de mesure et regroupent les nœuds redondants. Après un temps suffisant d'observation (*IntMeasure*), les CHs analysent les mesures collectées puis en se basant sur la proximité géographique des nœuds, ils déterminent les groupes de nœuds redondants.

Après l'identification des groupes redondants, un seul nœud représentant par groupe transmet à la fois et les autres nœuds arrêtent leurs transmissions. Tel que c'est illustré par la Figure 2, à la fin de la phase d'initialisation, MTMR fonctionne par super-trames périodiques. Les super-trames sont divisées en un ensemble de slots de temps qui sont alloués aux nœuds représentants pour envoyer leurs données au CH. Les autres slots sont assignés aux CHs pour envoyer les données à la station de base. Après chaque super-trame, le rôle du nœud représentant tourne entre les nœuds du groupe. L'ensemble des super-trames forme un round et MTMR assure que dans chaque round tous les nœuds envoient leurs données au moins une fois.

Il y a une proximité géographique si la distance entre les nœuds est inférieure au seuil donné d . Cette proximité se déduit directement de la distance euclidienne entre les nœuds.

On note x_i et y_i les coordonnées du $i^{\text{ème}}$ nœud, et on imagine que chaque nœud est un point $N_i = (x_i, y_i)$ de l'espace. La distance euclidienne entre deux nœuds N_i et N_j est par définition :

$$d_2(N_i, N_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

Pour qu'il y ait proximité géographique entre un ensemble de nœuds, il faut que :

$$d_2(N_i, N_j) \leq d + \varepsilon \quad (2)$$

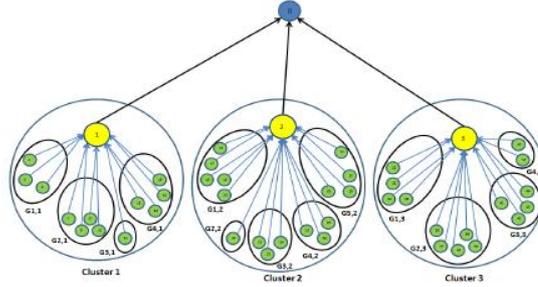


Figure 1. Réseau de capteurs après regroupement par proximité géographique

La proximité géographique est seulement vérifiée entre les nœuds capteurs qui reportent la même mesure durant l'intervalle de temps *IntMeasure*. Elle est donc vérifiée entre les capteurs qui sont susceptibles d'être redondants. La Figure 1 montre l'architecture du réseau après l'application de notre technique de regroupement.

Durant la phase d'observation des mesures (*IntMeasure*), chaque CH crée une table de valeurs (*ValTable*) où il va stocker la valeur de mesure reportée par chacun de ses capteurs membres. Cette table servira, par la suite, pour calculer la matrice de redondance (*reGroup*) comme détaillé dans l'algorithme 1. Chaque CH va ensuite vérifier la proximité géographique de tous les groupes de nœuds capteurs ayant reporté la même valeur de mesure (avec une certaine précision α) durant *IntMeasure*. Chaque CH va ensuite placer dans une même ligne de la matrice *reGroup* les identifiants de tous nœuds capteurs qui sont dans une même proximité. Après ce regroupement de proximité géographique, le CH demande à certains nœuds capteurs de s'endormir pendant la prochaine super-trame de telle sorte qu'un seul nœud par groupe transmettra à la fois durant la prochaine super-trame.

Ce processus est répété jusqu'à la fin du round qui correspond à la durée de vie des clusters comme montré dans la Figure 2. Il y aura à nouveau la formation des clusters ainsi que l'application de la technique de regroupement par proximité géographique.

Algorithme 1 : Procédure de calcul du regroupement

Require: *nbNodes, nbSent Measures*
Ensure: *reGroup[][]*

```

if (isClusterHead) then
  for (j ← 0; j < nbSentMeasures; j++) do
    for (i ← 0; i < nbNodes; i++) do
      valTable[j][i] ← valuei,j;
    end for
  end for
  {groups calculation}
  for (j ← 0; j < nbSentMeasures; j++) do
    f ← 0;
    x ← 0;
    for (i ← 0; i < nbNodes; i++) do
      value ← valTable[j][i];
      while (x ≠ nbNodes) do
        if (valTable[j][x] = value ± α and (d2(Ni, Nj) ≤ d + ε) then
          reGroup[f][i] ← idNodex;
          f ← f + 1;
          x ← x + 1;
        end if
      end while
    end for
  end for
end if
Return reGroup[][]

```

4. Évaluation de Performance

Pour évaluer les performances de MTMR, nous l'avons développé dans le simulateur réseau NS2 qui offre une possibilité d'intégrer le protocole LEACH. Tout d'abord, nous avons intégré l'implémentation de LEACH. Ensuite, pour développer notre technique de regroupement par proximité géographique, nous avons modifié LEACH selon la structure hiérarchique dans les commandes d'association de la couche MAC ainsi que la méthode de remplissage de la table de redondance au niveau de la couche physique.

Nous avons comparé notre solution avec LEACH. Pour cela, nous avons effectué des simulations avec les mêmes paramètres et métriques pour les deux approches. Nous nous sommes intéressés essentiellement à la consommation d'énergie des nœuds puisqu'elle constitue un critère primordial pour la détermination de la durée de vie d'un réseau de capteurs. Nous avons évalué le délai de bout en bout puisqu'il constitue un critère de performance pour beaucoup d'applications de monitoring dans les WSNs.

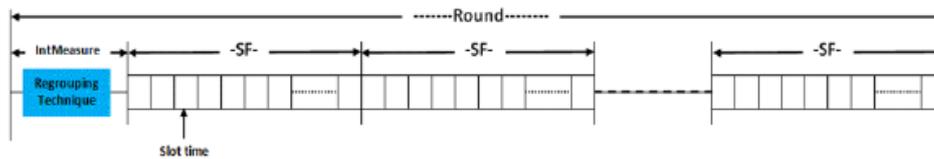


Figure 2. Représentation du round de la solution

4.1. Paramètres de Simulation

Les réseaux simulés sont de taille $100 \times 100 \text{ m}^2$. Le nombre maximal de nœuds est 1000. Les coordonnées de la station de base sont (50, 75). Les nœuds connaissent leur position et peuvent la communiquer à leur CH respectif.

Table I : Paramètres de simulation

Simulation area	$100 * 100 \text{ m}^2$
Transmission range	50 m
Sensing range	25 m
number of nodes	[100, 1000]
Packet length	55 bytes
Initial node energy	2 j
Transmission bit rate	250 kbs^{-1}
Transmit power	24.75 mW
Receive power	13.5 mW
Idle power	13.5 mW
Sleep power	$15 \mu\text{W}$

4.2. Métriques de performance

Les métriques de performance dans notre étude sont les suivantes :

- Consommation d'énergie : Pour évaluer la consommation d'énergie, nous avons récupéré les niveaux d'énergie au démarrage et les niveaux d'énergie courants des nœuds capteurs à partir des fichiers traces générés par les simulations.
- Délai de bout en bout : Pour évaluer le délai de bout en bout, nous avons récupéré le temps d'émission et de réception de chaque paquet à partir des fichiers traces générés.
- La durée de vie : Pour évaluer la durée de vie du réseau, nous avons récupéré le temps après lequel le niveau d'énergie d'un premier nœud est nul. Les niveaux d'énergie sont récupérés après chaque round.

Les résultats obtenus pour ces métriques sont présentés dans les figures 3, 4, 5 et 6 (avec 5% des nœuds comme CHs).

Amélioration de la consommation d'énergie fondée sur la redondance de mesure dans les WSNs

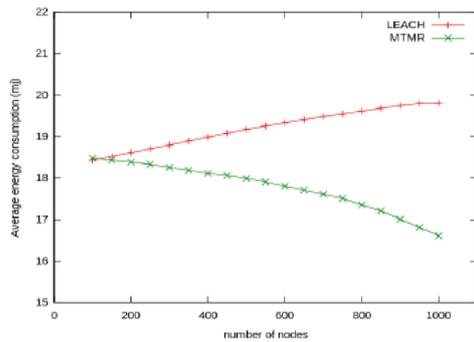


Figure 3. Consommation moyenne d'énergie de chaque CH durant un round

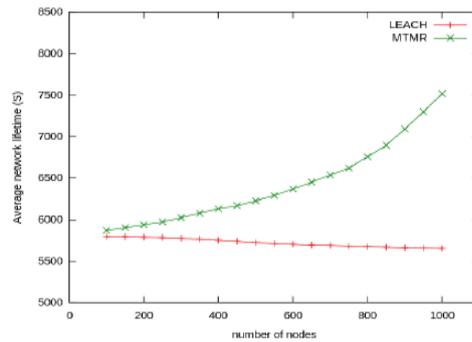


Figure 5. Durée de vie moyenne du réseau

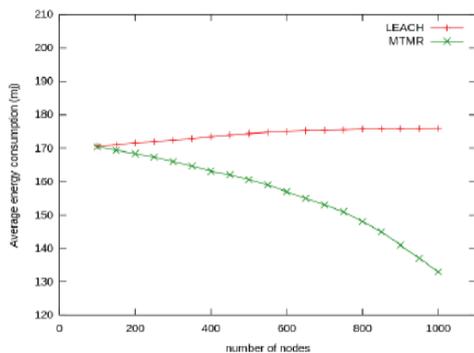


Figure 4. Consommation moyenne d'énergie de chaque nœud après 20 rounds

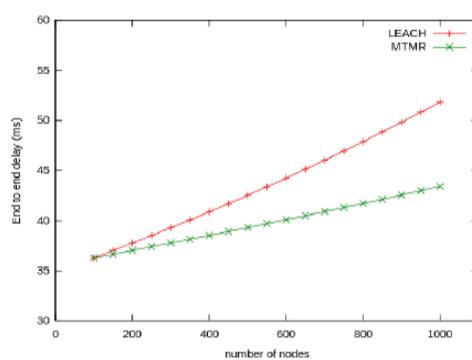


Figure 6. Délai de bout en bout

4.3. Comparaison

Comme l'illustre la Figure 3, nous remarquons que dans MTMR la consommation d'énergie des CHs diminue quand le nombre de nœuds augmente contrairement à LEACH. En effet, plus la densité du réseau augmente, plus la redondance de mesure augmente et moins les CHs reçoivent des données. Avec une forte densité, le nombre de groupes diminue dans les clusters. Ainsi, le nombre de communications diminue dans les clusters.

La Figure 4 montre la consommation moyenne d'énergie des nœuds durant 20 rounds de 25 secondes chacun. Dans nos simulations, la durée de 20 rounds permet à chaque nœud de prendre toutes les fonctions dans le réseau. Nous remarquons que dans MTMR la consommation d'énergie des nœuds diminue lorsque le nombre de nœuds augmente contrairement à LEACH. En effet, avec une forte densité, un nœud va communiquer moins avec son CH et les CHs reçoivent moins de paquets du fait de la redondance de mesure. Plus le nombre de nœuds augmente, plus les nœuds sont proches et moins il y a de groupes par proximité géographique dans les clusters. Donc, il y a moins de communications intra-clusters.

Comme le montre les courbes de la figure 5, nous remarquons que la durée de vie du réseau est presque stationnaire dans LEACH. Cependant, dans MTMR, cette durée de vie augmente avec le nombre de nœuds. La durée de vie du réseau est strictement liée à la consommation d'énergie des nœuds. Cette consommation d'énergie diminue considérablement quand le nombre de nœuds augmente (comme le montre la Figure 4).

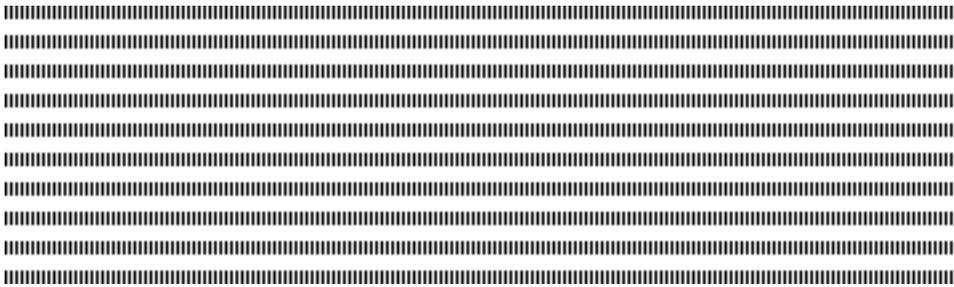
Nous remarquons d'après la Figure 6 que le délai de bout en bout augmente avec la densité du réseau. Cependant, cette augmentation est moins importante dans MTMR du fait de la prise en compte de la redondance de mesure. En fait, le CH dans MTMR attend un nombre $k \ll n$ de paquets avant de faire l'agrégation; où n est la taille du cluster.

5. Conclusion

Les réseaux de capteurs sans fil qui peuvent être denses dans une zone géographique, trouvent des applications dans de nombreux domaines. Dans ce contexte, il est nécessaire de mettre en place des approches pour une meilleure gestion de ces réseaux. Nous avons proposé MTMR (*Management Technique of the Measurement Redundancy*), une technique de regroupement de nœuds capteurs dans les clusters. Il introduit la notion de proximité géographique pour exploiter cette redondance. Cette technique améliore considérablement la consommation d'énergie et garantit un meilleur délai de bout en bout. Par conséquent, cette amélioration de la consommation d'énergie augmente la durée de vie du réseau. Les résultats des simulations montrent que notre approche permet de réduire considérablement la consommation d'énergie dans les réseaux de capteurs et augmente ainsi leur durée de vie. Il diminue également le délai de bout en bout.

6. Bibliographie

- [1] J. N. Al-Karaki, Raza Ul-Mustafa, and Ahmed E. Kamal. Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms. *Computer Networks*, 53(7):945 – 960, 2009.
- [2] J.N. Al-Karaki, R. Ul-Mustafa, and A.E. Kamal. Data aggregation in wireless sensor networks - exact and approximate algorithms. In *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on*, pages 241 – 245, 2004.
- [3] S. Bandyopadhyay and E.J. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pages 1713 – 1723, march-3 april 2003.
- [4] V. Geetha, PV Kallapur, and S. Tellajeera. Clustering in wireless sensor networks: Performance comparison of leach & leach-c protocols using ns2. *Procedia Technology*, 4:163–170, 2012.
- [5] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *Wireless Communications, IEEE Transactions on*, 1(4):660–670, October 2002.
- [6] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, page 10 pp. vol.2, jan. 2000.
- [7] T.T. Huynh and C.S. Hong. An energy* delay efficient multihop routing scheme for wireless sensor networks. *IEICE transactions on information and systems*, 89(5):1654–1661, 2006.
- [8] M. Lehsaini, H. Guyennet, and M. Feham. A novel clusterbased self-organization algorithm for wireless sensor networks. In *Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on*, pages 19 –26, may 2008.
- [9] M. Lehsaini, H. Guyennet, and M. Feham. An efficient clusterbased self-organisation algorithm for wireless sensor networks. *International Journal of Sensor Networks*, volume 7:85–94, March. 2010.
- [10] S. Lindsey, C. S. Raghavendra, and K. M. Sivalingam. Data gathering in sensor networks using the energy delay metric. In *Proceedings of the 15th International Parallel & Distributed Processing Symposium, IPDPS '01*, pages 188–, Washington, DC, USA, 2001. IEEE Computer Society.
- [11] S. Lindsey and C.S. Raghavendra. Pegasus: Power-efficient gathering in sensor information systems. In *Aerospace Conference Proceedings, 2002. IEEE*, volume 3, pages 1125 – 1130, 2002.
- [12] A. Manjeshwar and D.P. Agrawal. Teen: a routing protocol for enhanced efficiency in wireless sensor networks. In *Parallel and Distributed Processing Symposium., Proceedings 15th International*, pages 2009 –2015, april 2001.
- [13] A. Manjeshwar and D.P. Agrawal. Apteem: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In *Parallel and Distributed Processing Symposium., Proceedings International, IPDPS 2002, Abstracts and CD-ROM*, pages 195 –202, 2002.
- [14] M. Perillo and W. Heinzelman. Wireless sensor network protocols. *Algorithms and Protocols for Wireless and Mobile Networks, Eds. A. Boukerche et al., CRC Hall Publishers*, 2004.
- [15] A. Savvides, C.C. Han, and M.B. Strivastava. Dynamic finegrained localization in ad-hoc networks of sensors. In *International Conference on Mobile Computing and Networking: Proceedings of the 7th annual international conference on Mobile computing and networking*, vol 2001, pages 166–179.
- [16] O. Younis and S. Fahmy. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *Mobile Computing, IEEE Transactions on*, 3(4):366 – 379, oct.- dec. 2004.
- [17] Isabel Dietrich and Falko Dressler. On the lifetime of wireless sensor networks. *ACM Trans. Sen. Netw.*, 5(1):5:1–5:39, February 2009.



Factorisation multi-biais pour de meilleures recommandations

Modou Gueye^{1,2} — Talel Abdesslem¹ — Hubert Naacke³

¹ Institut Telecom, Telecom ParisTech
46, rue Barrault 75013, Paris, France
prénom.nom@telecom-paristech.fr

² LID, Université Cheikh Anta Diop
BP. 16432, Dakar-Fann, Sénégal
gmodou@ucad.sn

³ LIP6, UPMC Sorbonne Universités
4, place Jussieu 75005, Paris, France
hubert.naacke@lip6.fr

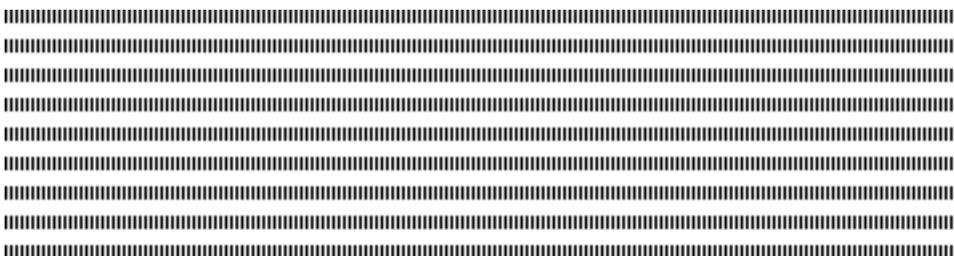


RÉSUMÉ. La factorisation biaisée de matrices offre une grande qualité de prédiction pour les systèmes de recommandation. Mais elle attribue un comportement uniforme à tous les utilisateurs et produits. Ce papier propose une amélioration de la factorisation biaisée de matrices. Notre proposition utilise des groupes de produits similaires. Elle attribue à chaque utilisateur un biais pour chaque groupe de produits. Nos expérimentations sur des jeux de données réels montrent l'efficacité de notre approche.

ABSTRACT. It is today accepted that adding biases to matrix factorization models allows a high quality of rating prediction in recommender systems. However using a single bias per user limits a user to have an uniform behaviour. In this paper, aiming at improving the accuracy of recommender systems, we propose a cluster-based matrix factorization technique. We use finer-grained user biases by clustering similar items into groups, and allocating in these groups a bias to each user. The experiments we did on large datasets demonstrated the efficiency of our approach.

MOTS-CLÉS : Recommandation en ligne, Factorisation de matrices

KEYWORDS : Online recommendation, Matrix factorization



1. Introduction

L'objectif des systèmes de recommandation est de déterminer, parmi une grande quantité de produits, lesquels intéresseront le plus un utilisateur donné. Plus les produits recommandés intéressent effectivement l'utilisateur, plus la qualité de la recommandation est grande. La recommandation a une valeur commerciale capitale pour tout type de e-commerce et concentre aujourd'hui beaucoup d'attention afin d'améliorer sa qualité [2, 12].

La factorisation de matrices est une technique de filtrage collaboratif apportant une qualité très satisfaisante [8, 10, 11, 13, 15]. Elle consiste à construire des profils caractérisant les utilisateurs et les produits, au moyen de vecteurs de facteurs. Ces profils sont déduits des notes que les utilisateurs attribuent aux produits. Ainsi, il est possible d'estimer l'intérêt d'un utilisateur pour un produit en combinant le profil de l'utilisateur avec celui du produit (ex. : produit scalaire). Puis, les produits avec les plus grandes estimations sont recommandés.

Il est connu que certains utilisateurs ont tendance à donner des notes plus élevées ou plus faibles que les autres utilisateurs. Et certains produits aussi sont plus ou moins appréciés que les autres. De ce fait, la prise en compte de ces tendances dans la factorisation est l'une des extensions améliorant la qualité des recommandations [10, 11, 15].

Nos travaux visent à améliorer la qualité des recommandations. Nous proposons une factorisation multi-biais. Nous partons de l'observation que les utilisateurs ont tendance à noter différemment des produits de différents groupes et attribuons à chaque utilisateur une liste de biais décrivant ses tendances. Les expérimentations que nous avons effectuées sur le jeu de données de Netflix et le plus grand de MovieLens confirment que notre approche améliore de façon effective la qualité des recommandations [3, 16].

Ce papier est organisé comme suit. A la section 2, nous introduisons des notions préliminaires sur la recommandation. Puis, à la section 3, nous présentons notre solution de factorisation basée sur des groupes de similarité. De même, nos résultats expérimentaux sont présentés à la section 4. Enfin, à la section 5, nous exposons des travaux connexes, avant de conclure.

2. Préliminaires

Le but d'un système de recommandation est de découvrir les intérêts d'un utilisateur afin de lui proposer des produits susceptibles de l'intéresser. La plupart du temps, l'intérêt d'un utilisateur sur un produit est donné sous forme de valeurs numériques d'une plage donnée (ex. : 1 à 5) et il ne note un produit qu'une fois. Plus la valeur donnée est élevée, plus l'utilisateur s'intéresse au produit.

Si on considère un ensemble d'utilisateurs U , un ensemble de produits I et une liste de notes (u, i, r_{ui}, t_{ui}) où chaque valeur r_{ui} représente l'intérêt de l'utilisateur u pour le

produit i - t_{ui} étant le moment où la note a été soumise, la recommandation consiste à prédire les futures notes des utilisateurs telles que l'écart entre une note prédite et celle réellement donnée ultérieurement, soit le plus petit possible. Cela permet de proposer à l'utilisateur les produits présentant les plus grandes valeurs de prédiction. Ainsi, la qualité d'un système de recommandation peut-être rattachée à la précision de ses prédictions. En pratique, pour estimer cette précision (c-à-d les écarts), l'ensemble des notes existantes est subdivisé en deux parties : la plus grande pour l'apprentissage et la seconde pour l'évaluation. La mesure appelée RMSE est l'une des plus utilisées pour l'évaluation. RMSE est la racine carré de la moyenne des carrés des écarts [5, 13, 16]. Nous l'utilisons pour la suite.

$$RMSE = \sqrt{\frac{1}{n} \sum_{u,i} (r_{ui} - \hat{r}_{ui})^2} \quad (1)$$

n représente le nombre total de notes à prédire. Plus petit est le RMSE, meilleures sont les prédictions.

2.1. factorisation de matrices

Les systèmes de recommandation utilisant la factorisation de matrices représentent les notes des utilisateurs dans une matrice R creuse. Les colonnes représentent les utilisateurs et les lignes les produits. Ainsi la note $r_{ui} \in R$ est celle donnée par l'utilisateur u au produit i . R est généralement très creuse [16]. L'objectif de la factorisation est de prédire les valeurs manquantes dans R . Dans sa forme basique (FM basique), elle cherche à approximer R comme le produit de deux autres matrices

$$R = P \cdot Q \quad (2)$$

Les deux matrices P et Q contiennent respectivement les vecteurs de facteurs des utilisateurs et ceux des produits. Ce sont les matrices de facteurs. Pour prédire la note \hat{r}_{ui} que l'utilisateur u donnerait au produit i , il suffit simplement d'appliquer la formule

$$\hat{r}_{ui} = p_u \cdot q_i^T \quad (3)$$

p_u et q_i étant respectivement les vecteurs de facteurs de l'utilisateur u et du produit i dans P et Q .

Le processus d'apprentissage qu'effectue la factorisation détermine les valeurs dans P et Q telles qu'on s'approche le plus des notes r_{ui} existantes dans R . Il utilise une descente de gradient stochastique (DGS) qui calcule un minimum local tel que la somme des erreurs (c-à-d des écarts), $e_{ui} \stackrel{def}{=} r_{ui} - \hat{r}_{ui}$ entre les notes prédites \hat{r}_{ui} et celles réelles r_{ui} données par les utilisateurs, soit la plus faible possible. DGS minimise la somme

des erreurs quadratiques $\sum_{ui} e_{ui}^2$ en ajustant les facteurs dans P et Q jusqu'à ce que cette somme ne diminue plus :

$$\begin{aligned} p_{uk} &\leftarrow p_{uk} + \lambda \cdot (2 \cdot e_{ui} \cdot q_{ki} - \beta \cdot p_{uk}) \\ q_{ki} &\leftarrow q_{ki} + \lambda \cdot (2 \cdot e_{ui} \cdot p_{uk} - \beta \cdot q_{ki}) \end{aligned} \quad (4)$$

Ceci permet de diminuer les erreurs et par conséquent d'avoir une meilleure approximation des notes réelles. Le paramètre λ introduit dans l'ajustement des facteurs représente un taux d'apprentissage. β est un paramètre de régularisation.

Après cette phase, les prédictions \hat{r}_{ui} sont calculées comme les produits $p_u \cdot q_i^T$. Un tri est effectué par la suite pour trouver les produits les plus intéressants (ceux avec les plus grandes notes de prédiction) et les recommander à l'utilisateur concerné.

2.2. Factorisation biaisée de matrices (FBM)

Plusieurs améliorations possibles de la factorisation présentée ci-dessus ont été proposées dans la littérature.

L'une d'elle suppose que la plupart des variations observées sur les notes des utilisateurs sont dues principalement à des effets associés soit aux utilisateurs, soit aux produits [10, 11, 15]. Autrement dit, certains utilisateurs ont tendance à donner des notes plus élevées ou plus faibles que les autres utilisateurs. Et certains produits aussi sont plus ou moins appréciés que les autres. La factorisation basique (FM basique) présentée précédemment ne prend pas en compte ces tendances. La factorisation biaisée de matrices (FBM) introduit des biais pour tenir en compte ces variations de notation. Les biais reflètent les tendances des utilisateurs et des produits. On a la formule de prédiction suivante :

$$\hat{r}_{ui} = p_u \cdot q_i^T + \mu + b_u + b_i \quad (5)$$

où μ dénote la moyenne de toutes les notes confondues, b_u et b_i sont respectivement le biais de l'utilisateur et celui du produit (i.e, la tendance de l'utilisateur et la perception du produit par rapport à la moyenne). Une bonne approximation de ces biais est cruciale pour avoir des prédictions de bonne qualité [8, 11]. Ainsi, ils doivent être ajustés durant la phase d'apprentissage en utilisant

$$\begin{aligned} b_i &\leftarrow b_i + \lambda \cdot (2 \cdot e_{ui} - \gamma \cdot b_i) \\ b_u &\leftarrow b_u + \lambda \cdot (2 \cdot e_{ui} - \gamma \cdot b_u) \end{aligned} \quad (6)$$

γ est un paramètre de régularisation. Il joue un rôle comparable à β dans l'équation 4.

3. Factorisation multi-biais de matrices (FMBM)

Beaucoup d'utilisateurs tendent à sur-appécier ou sous-appécier les produits qu'ils notent. Une manière simple de quantifier cette tendance est d'assigner un biais global à

chaque utilisateur comme avec la FBM [10, 11, 14]. Cependant, la tendance d'un utilisateur n'est généralement pas uniforme : elle peut changer d'un groupe de produits à un autre. Pour certains groupes de produits, un utilisateur peut avoir tendance à noter comme tout le monde alors qu'il surestime ou sous-estime d'autres groupes par manque d'objectivité. Cette tendance devient uniforme pour un ensemble de produits similaires ; ce que nous avons formalisé dans nos travaux [4] à travers l'équation bornant la variance de notation des utilisateurs par la dissimilarité des produits notés :

$$0 \leq \sum_{u \in U} Var_u \leq \left(\sum_{(i,j) \in I^2} dissim_{ij} \right)^2 \quad (7)$$

Pour prendre en compte cette diversité de notation, nous attribuons un biais local $\delta_u^{c(i)}$ à chaque utilisateur u pour chaque groupe C de produits jugés similaires ou proches. Cette multitude de biais par utilisateur permet d'avoir un modèle plus raffiné et vise à une meilleure qualité de recommandation.

Notre formule de prédiction est la suivante

$$\hat{r}_{ui} = p_u \cdot q_i^T + \mu^{c(i)} + \delta_u^{c(i)} + b_u + b_i \quad (8)$$

où $c(i)$ représente le groupe auquel appartient le produit i .

Nous affinons les biais locaux $\delta_u^{c(i)}$ durant la phase d'apprentissage en utilisant :

$$\delta_u^{c(i)} \leftarrow \delta_u^{c(i)} + \lambda \cdot (2 \cdot e_{ui} - \gamma \cdot \delta_u^{c(i)}) \quad (9)$$

L'algorithme 1 détaille les étapes de la génération de modèle par la FMBM. A la ligne 1, un clustering est effectué afin de déterminer les groupes de produits. La ligne 2 calcule les valeurs initiales des biais des produits, celles des biais globaux et locaux des utilisateurs. La ligne 3 initialise les matrices de facteurs P et Q par des valeurs aléatoires faibles. Les lignes 4 à 11 constituent la partie principale de l'apprentissage. A chaque itération (lignes 5 à 10), l'erreur de prédiction e_{ui} est calculée pour chaque note. Puis les matrices de facteurs, les biais (globaux et locaux) sont ajustés en conséquence (lignes 7 à 11) en utilisant les équations 4, 6 et 9. La ligne 13 calcule la somme globale des erreurs quadratiques. L'apprentissage s'arrête lorsque cette somme ne décroît plus ou lorsqu'un nombre spécifié d'itérations a été atteint.

4. Expérimentation

Nous avons utilisé les jeux de données de Netflix et MovieLens dont les tailles respectives sont de 100M et 10M de notes [3, 16]. Ces jeux de données sont très utilisés dans la littérature [13].

Notre implémentation a été faite sous C++ et nous avons effectué un calibrage préliminaire pour fixer nos paramètres : $\lambda = 0,001$, $\beta = 0,02$, $\gamma = 0,05$, $N_c = 3$. Ces valeurs

Algorithm 1: Algorithme de la FMBM.**Data:** N_c : nombre de groupes, K : nombre de facteurs, \mathbf{R} , λ , β et γ **Result:** P , Q , $\mu = \{\mu^C\}$, b_i , b_u et $\{\delta_u^C\}$, $C \in C_1, C_2, \dots, C_{N_c}$

```

1 Créer les groupes  $C_1, C_2, \dots, C_{N_c}$  à partir de  $\mathbf{R}$ ;
2 Pour chaque produit  $i$ , utilisateur  $u$ , calculer  $b_i$ ,  $b_u$  et  $\{\delta_u^C\}$ ,  $C \in C_1, C_2, \dots, C_{N_c}$ ;
3 Initialiser les matrices  $P$  et  $Q$ ;
4 repeat
5   foreach  $r_{ui} \in \mathbf{R}$  do
6     Calculer  $e_{ui}$ ;
7     for  $k \leftarrow 1$  to  $K$  do
8       Ajuster  $p_{uk} \in P$ ,  $q_{ki} \in Q$ ;
9     end
10    Ajuster  $b_i$ ,  $b_u$  et  $\delta_u^{C(i)}$ ;
11  end
12  Calculer la somme globale des erreurs  $\sum_{r_{ui}>0} e_{ui}^2$ ;
13 until condition d'arrêt;
14 return  $P$ ,  $Q$ ,  $\mu = \{\mu^C\}$ ,  $b_i$ ,  $b_u$ ,  $\{\delta_u^C\}$ ,  $C \in C_1, C_2, \dots, C_{N_c}$ 

```

de λ , β , et γ sont proches de celles dans [11]. Nous avons limité le nombre d'itérations à 120 au plus et utilisons 40 facteurs pour P et Q . Pour le clustering des produits, nous avons effectué une FM basique de 30 itérations suivie d'un K-Means avec la distance euclidienne.

Pour notre expérimentation nous avons trié les données par date croissante puis nous les avons séparées en deux parties : la première pour l'apprentissage (98% des notes, les plus anciennes) et le reste (les 2% de notes les plus récentes) pour le test. Ainsi, notre jeu de test pour Netflix contient 1,88M de notes, une taille proche de celle du Netflix Prize dont le jeu de test contenait 1,4M de notes [16]. La table 1 rapporte les différentes erreurs RMSE obtenues avec les trois modèles FM basique, FBM et FMBM. FMBM a les plus petites erreurs et dépasse donc les deux autres en termes de qualité. Elle atteint 1,12% de qualité de plus que la FBM (Notons à ce point qu'une amélioration aussi petite que 1% se traduit par une différence significative dans la plupart des "top-K" produits recommandés aux utilisateurs [1, 7]).

Nous avons effectué plusieurs autres expérimentations dans le cadre de la validation de notre approche. Nous n'avons pu les mettre dans ce papier pour des raisons de limitation de taille. Elles sont disponible dans [4].

5. Travaux connexes

Ce papier propose une amélioration de la FBM avec l'utilisation de plusieurs biais. Dans la littérature, on trouve beaucoup de propositions allant dans ce sens. Mais on peut

Jeu de données	FM basique	FBM	FMBM
MovieLens	0,7743	0,7608	0,7578
Netflix	0,9599	0,9312	0,9208

Tableau 1. RMSEs des trois modèles (FM basique, FBM et FMBM)

les catégoriser brièvement en deux groupes. Les premiers prennent en considération l'aspect temporel. Ils utilisent des sessions de temps pour suivre l'évolution des tendances des utilisateurs et des produits [9, 10, 14]. Ils introduisent ainsi des biais par session de temps. Le second groupe quant à lui prend en compte les types de produits selon leurs attributs [6] et attribue plusieurs biais aux utilisateurs. Notre travail se situe dans ce cadre. Sa particularité réside sur le fait qu'on se base sur un formalisme bien défini sur l'apport de l'utilisation de plusieurs biais par utilisateur [4]. De plus, notre proposition n'a pas besoin d'informations supplémentaires sur les produits pour les classer. Elle se base uniquement sur les notes qu'ils reçoivent.

6. Conclusion

Nous nous sommes intéressés à l'amélioration de la qualité de la factorisation dans le cadre de la recommandation. Notre solution introduit des biais personnalisés pour les utilisateurs afin de mieux capturer leurs subjectivités selon des groupes de produits. Nous regroupons les produits à partir des notes qu'ils reçoivent.

Nos résultats expérimentaux montrent que notre solution surpasse celle utilisant un seul biais global par utilisateur et celle n'en ayant pas.

7. Bibliographie

- [1] Gideon Dror, Noam Koenigstein, Yehuda Koren, and Markus Weimer. The yahoo! music dataset and kdd-cup'11. In *Proceedings of KDD Cup 2011*, 2011.
- [2] Daniel M. Fleder and Kartik Hosanagar. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce, EC '07*, pages 192–199, New York, NY, USA, 2007. ACM.
- [3] <http://www.grouplens.org/node/73>.
- [4] Modou Gueye, Talel Abdesslem, and Hubert Naacke. Dynamic recommender system : using cluster-based biases to improve the accuracy of the predictions. Technical report, arXiv.org, 2012.
- [5] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22 :5–53, 2004.

- [6] Noam Koenigstein, Gideon Dror, and Yehuda Koren. Yahoo ! music recommendations : modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of the 5th ACM conference on Recommender systems, RecSys '11*, pages 165–172, New York, NY, USA, 2011. ACM.
- [7] Yehuda Koren. How useful is a lower rmse ?, 2007. Netflix Prize Forum.
- [8] Yehuda Koren. The bellkor solution to the netflix grand prize, 2009.
- [9] Yehuda Koren. Collaborative filtering with temporal dynamics. *Commun. ACM*, 53(4) :89–97, 2010.
- [10] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42 :30–37, 2009.
- [11] Arkadiusz Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proc. KDD Cup Workshop at SIGKDD'07, 13th ACM Int. Conf. on Knowledge Discovery and Data Mining*, pages 39–42, 2007.
- [12] J. Ben Schafer, Joseph Konstan, and John Riedi. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce, EC '99*, pages 158–166, New York, NY, USA, 1999. ACM.
- [13] Xiaoyuan Su and Taghi M. Khoshgoftaar. A survey of collaborative filtering techniques. *Adv. in Artif. Intell.*, 2009 :4 :2–4 :2, 2009.
- [14] Gábor Takács, István Pilászy, Botyán Németh, and Domonkos Tikk. Investigation of various matrix factorization methods for large recommender systems. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition, NETFLIX '08*, pages 6 :1–6 :8, New York, NY, USA, 2008. ACM.
- [15] Gábor Takács, István Pilászy, Botyán Németh, and Domonkos Tikk. Scalable collaborative filtering approaches for large recommender systems. *J. Mach. Learn. Res.*, 10 :623–656, 2009.
- [16] http://en.wikipedia.org/wiki/netflix_prize.

1. Introduction

Le problème de l'exploration de graphes consiste, pour une entité mobile (aussi appelée agent), à explorer tous les sommets (ou toutes les arêtes) d'un graphe a priori inconnu. Ce problème étant l'un des plus classiques dans le cadre du calcul par agent mobile, il a reçu beaucoup d'intérêt jusqu'ici. La complexité en temps, en espace, ou l'impact d'une connaissance a priori ont été largement étudiés au cours des quarante dernières années (voir par exemple, [3, 10, 11]). Cependant, la grande majorité de ces travaux concerne les graphes statiques, tandis que les nouvelles générations d'environnements inter-connectés tendent à être extrêmement dynamiques. Dans cet article, nous étudions le problème de l'exploration de graphes dans un modèle de réseaux dynamiques, à savoir le modèle des graphes dynamiques T -intervalle-connexes ($T \geq 1$) (voir la section 3 pour les définitions). Le lecteur intéressé trouvera dans [1] un récapitulatif très complet des différents modèles de graphes dynamiques existants.

Un graphe dynamique est T -intervalle-connexe ($T \geq 1$) si pour chaque fenêtre de T unités de temps, il existe un sous-graphe couvrant connexe stable. Cette propriété de stabilité de graphes dynamiques a été introduite par Kuhn, Lynch et Oshman [9] (STOC 2010). Elle permet d'accélérer le temps de calcul de certaines fonctions calculables en distribué.

Dans cet article, nous étudions la complexité en temps de l'exploration des graphes dynamiques T -intervalle-connexes qui ont comme graphe sous-jacent un anneau de taille n . Nous allons considérer le cas où l'agent connaît la dynamique du graphe avant d'entamer l'exploration, c'est à dire qu'il connaît les temps d'apparitions et de disparitions des arêtes du graphe sous-jacent, et le cas où l'agent ne connaît pas la dynamique du graphe. Dans ce dernier cas, nous allons supposer que le graphe dynamique est δ -récurrent. En effet, sans hypothèse supplémentaire, l'exploration des graphes dynamiques T -intervalle-connexes est impossible [9]. Un graphe dynamique est δ -récurrent si chaque arête du graphe sous-jacent apparaît au moins une fois toutes les δ unités de temps.

Nous allons montrer que, si l'agent connaît la dynamique du graphe, $2n - T - \Theta(1)$ unités de temps sont nécessaires et suffisantes pour explorer les graphes dynamiques T -intervalle-connexes qui ont comme graphe sous-jacent un anneau de taille n . Si l'agent ne connaît pas la dynamique des graphes, nous supposons que les graphes dynamiques sont δ -récurrents et nous montrons que le temps d'exploration dans le pire cas est $\frac{n}{\max\{1, T-1\}}(\delta - 1) + n \pm \Theta(\delta)$ unités de temps.

2. Etat de l'art

Ces dernières années, un travail de recherche important a été fait dans les graphes dynamiques, notamment dans l'exploration et la cartographie de ces environnements très dynamiques par des agents mobiles. Motivés par l'exploration robotisée du Web, Cooper et Frieze [2] ont étudié en 2003 la question du temps de couverture minimal d'un graphe qui évolue au cours du temps. Ils ont considéré un modèle particulier de graphes du web et montrent que si après chaque nombre constant de pas de la marche un nouveau sommet apparaît et est raccordé au graphe, une marche aléatoire probabiliste sur le graphe manque de visiter une fraction constante de sommets. Izumi et al [8] étudient le problème de rendez-vous dans un graphe dynamique modélisé par un graphe évolutif. Pour tout

graphe G , un graphe évolutif de G est une séquence de sous-graphes de G [4]. Flocchini et al. [5] introduisent un modèle de graphe très dynamique "subway model". Puis ils étudient le modèle avec des trous noirs. Pour cela, ils considèrent qu'il existe des agents mobiles au nombre de k qui se déplacent avec des transporteurs pour faire la carte du graphe. Casteigts et al. [1] intègrent dans un cadre unifié une vaste collection de concepts, de formalismes et de résultats obtenus dans la littérature des graphes variables dans le temps. Flocchini et al. [6] considèrent une famille de graphes variables, les graphes périodiquement variables, et montrent que la connaissance d'une borne supérieure sur la plus grande période ou du nombre n de sommets du graphe est nécessaire pour qu'un agent puisse explorer le graphe. Ilcinkas et Wade [7] considèrent cette même famille de graphes variables et montrent l'importance d'attendre aux sommets pour l'exploration de cette famille de graphes dynamiques. En 2010 Kuhn et al. [9] étudient les calculs distribués dans les graphes dynamiques. Pour cela, ils introduisent une propriété de stabilité de graphes dynamiques nommée T-intervalle-connexité. Puis ils montrent que si le graphe dynamique est 1-intervalle connexe, les sommets peuvent calculer toute fonction calculable en $O(n^2)$ rondes et que ce temps de calcul peut être accéléré par un facteur T si le graphe dynamique est T-intervalle-connexe.

3. Définitions et modèles

Dans cette section, nous allons donner le modèle de graphe dynamique que nous étudions, qui est le même que dans [9], et quelques définitions sur le modèle.

Définition 1 (*Grappe dynamique*) Nous considérons les réseaux dynamiques modélisés par le graphe dynamique $G = (V, E)$ où V est un ensemble de n sommets statiques, et $E : \mathbb{N} \rightarrow \{\{u, v\} | u, v \in V\}$ est une fonction qui associe au nombre entier $r \in \mathbb{N}$ un ensemble d'arêtes non orientées $E(r)$.

Définition 2 (*Grappe sous-jacent*) Le graphe $U(G) = (V, \bigcup_{r=0}^{\infty} E(r))$ est dit graphe sous-jacent du graphe dynamique. Il représente le graphe dynamique avec toutes les arêtes ayant existé, à un moment ou à un autre.

Nous considérons les graphes dynamiques $G(V, E)$ tel que $U(G)$ est un anneau de taille n noté A_n . Pour tout i et j , une arête entre s_i et s_{i+1} appartenant à l'anneau sera noté par $\{s_i, s_{i+1}\}$ et un chemin de s_i à s_j sera noté par $(s_i, s_{i+1}, \dots, s_j)$. Les indices sont pris modulo n .

Définition 3 (*Grappe dynamique T-intervalle-connexe*) Nous disons qu'un graphe dynamique est T-intervalle-connexe pour $T \geq 1$, si pour tout $r \in \mathbb{N}$ le graphe statique $G_{r,T} := (V, \bigcap_{i=r}^{r+T-1} E(i))$ est connexe.

Définition 4 (*Grappe dynamique δ -récurrent*) Un graphe dynamique est δ -récurrent si toute arête du graphe sous-jacent apparaît au moins une fois toutes les δ unités de temps.

Une entité externe, appelée *agent*, opère sur ces graphes dynamiques. Il peut traverser, en une unité de temps, au plus une arête. Nous disons que l'agent explore le graphe dynamique si et seulement si il peut visiter l'ensemble des sommets en un temps fini et se mettre dans un état terminal. Cet état terminal exprime le fait que l'agent détecte que l'exploration est terminée.

4. Bornes inférieures

Dans cette section, nous allons donner des bornes sur le temps nécessaire pour explorer un graphe dynamique T-intervalle-connexe qui a comme graphe sous-jacent A_n . Pour cela nous allons supposer en premier lieu que l'agent connaît la dynamique du graphe et nous allons finir par le cas où l'agent ne connaît pas la dynamique du graphe.

4.1. Le cas où l'agent connaît la dynamique du graphe

Nous supposons que l'agent connaît les temps d'apparitions et de disparitions des arêtes du graphe sous-jacent.

Théorème 1 *Pour tout $n \geq 3$ et $T \geq 1$, il existe un graphe dynamique $G_{n,T}$ T-intervalle-connexe qui a comme graphe sous-jacent A_n tel que tout agent doit faire au moins $2n - T - 2$ unités de temps pour l'explorer.*

Preuve Soit $n \geq 3$ et $T \geq 2$. Sans perte de généralité, supposons que l'agent est sur s_0 au temps $t = 0$. On construit $G_{n,T}$ de la manière suivante. L'arête $\{s_0, s_1\}$ respectivement $\{s_{T-1}, s_T\}$ est déconnectée aux temps $t \in [0, n - 2T + 1[$ respectivement $t \in [n - T, 2n[$. Toutes les autres arêtes de l'anneau sont toujours présentes. $G_{n,T}$ est T-intervalle-connexe. Pour explorer le graphe dynamique, tout agent (algorithme) doit visiter l'ensemble des sommets de A_n notamment les sommets s_{T-1} et s_T .

– Si $T \geq n - 1$ alors $2n - T - 2 \leq n - 1$ donc le résultat est trivial. Car pour explorer un anneau de taille n tout agent doit faire au moins $n - 1$ unités de temps.

– Si $1 < T < n - 1$

- supposons que l'agent visite s_{T-1} avant s_T . Pour visiter s_{T-1} sans passer par le sommet s_T , l'agent doit forcément passer par l'arête $\{s_0, s_1\}$. Par construction, l'arête $\{s_0, s_1\}$ n'est présente qu'après $n - 2T + 1$ unités de temps et que le chemin $(s_0, s_1, \dots, s_{T-1})$ est de longueur $T - 1$. Donc l'agent va payer au moins $n - T$ unités de temps pour visiter s_{T-1} pour la première fois sans passer par le sommet s_T . Comme l'arête $\{s_{T-1}, s_T\}$ est absente pour les temps $t \in [n - T, 2n[$, pour visiter le sommet s_T après avoir visité s_{T-1} , l'agent a besoin au moins de $2n - T - 1$ unités de temps.

- supposons que l'agent visite s_T avant s_{T-1} . Pour visiter s_T sans passer par le sommet s_{T-1} , l'agent doit forcément passer par le chemin $(s_T, s_{T+1}, \dots, s_0)$. Ce chemin est de longueur $n - T$. Donc l'agent doit payer au moins $n - T$ unités de temps pour visiter le sommet s_T pour la première fois sans passer par le sommet s_{T-1} . Comme l'arête $\{s_{T-1}, s_T\}$ est absente pour les temps $t \in [n - T, 2n[$, pour visiter le sommet s_{T-1} après avoir visité s_T , l'agent a besoin au moins de $2n - T - 1$ unités de temps.

– Si $T = 1$ en construisant $G_{n,T}$ de la manière suivante, l'arête $\{s_0, s_1\}$ respectivement $\{s_1, s_2\}$ est déconnectée aux temps $t \in [0, n - 3[$ respectivement $t \in [n - 2, 2n[$, et en raisonnant de la même façon que le cas précédant, nous pouvons prouver que la complexité est de $2n - 3$ unités de temps.

□

4.2. Le cas où l'agent ne connaît pas la dynamique du graphe

Ici nous supposons que l'agent ne connaît pas les temps d'apparitions et de disparitions des arêtes du graphe sous-jacent.

Théorème 2 Pour tout $n \geq 3$, $T \geq 1$, et $\delta \geq 1$, il existe un graphe dynamique $G_{n,T,\delta}$ T -intervalle-connexe et δ -récurrent de taille n qui a comme graphe sous-jacent l'anneau A_n tel que tout agent doit faire au moins $\lfloor \frac{n-1}{\max\{1, T-1\}} \rfloor (\delta - 1) + n - 1$ unités de temps pour l'explorer.

Ce résultat reste vrai même si l'agent connaît n et δ .

Preuve Soit $n \geq 3$, $T \geq 1$ et $\delta \geq 1$. L'adversaire construit le graphe dynamique $G_{n,T,\delta}$ comme suit.

- 1) L'adversaire met le temps d'exploration de l'agent à zéro.
 - 2) Si l'agent arrive sur un sommet en faisant T unités de temps, l'adversaire déconnecte l'arête sortante.
 - 3) Si l'agent décide d'aller au sommet précédant, l'adversaire remet l'arête et la retire si l'agent revient sur le même sommet avant $T + \delta$ unités de temps.
 - 4) Si l'agent décide de rester sur le sommet, l'adversaire remet l'arête au bout de δ unités de temps à partir du moment où l'arête a été déconnectée. Puis il recommence à 1).
- On peut bien voir qu'avec cette construction $G_{n,T,\delta}$ est T -intervalle-connexe et δ -récurrent. A chaque fois que l'agent visite $\max\{1, T - 1\}$ nouveaux sommets, l'adversaire l'oblige à attendre $(\delta-1)$ -unités de temps avant de pouvoir visiter un nouveau sommet. Comme le graphe dynamique a n sommets, alors l'agent va payer au moins $\lfloor \frac{n-1}{\max\{1, T-1\}} \rfloor (\delta - 1)$ unités de temps d'attente et $n - 1$ unités de temps d'exploration. Donc l'agent va payer au moins $\lfloor \frac{n-1}{\max\{1, T-1\}} \rfloor (\delta - 1) + n - 1$ unités de temps pour explorer le graphe. \square

5. Bornes supérieures

Dans la première partie de l'article, nous avons fourni des bornes inférieures sur la complexité en temps de l'exploration des graphes dynamiques T -intervalle-connexes qui ont comme graphe sous-jacent l'anneau A_n . Nous allons maintenant prouver que tous ces résultats sont optimaux (ou asymptotiquement optimaux). Nos bornes supérieures, dans les deux cas étudiés, utilisent comme base le même algorithme décrit ci-dessous.

Algorithme EXPLORE-ANNEAU (dir)

```

A chaque unité de temps faire
if l'arête en direction de dir est présente then
  la prendre
else
  attendre
end if

```

dir est la direction que va suivre l'agent exécutant l'algorithme depuis un sommet.

5.1. Le cas où l'agent connaît la dynamique du graphe

La connaissance de la dynamique du graphe est suffisante pour explorer tous les graphes dynamiques T -intervalle-connexes qui ont A_n comme graphe sous-jacent.

Théorème 3 Pour tout $n \geq 3$, $1 \leq T \leq \lfloor \frac{n-1}{2} \rfloor$ et $G_{n,T}$ un graphe dynamique T -intervalle-connexe de taille n qui a comme graphe sous-jacent A_n . Il existe un agent (algorithme) A capable d'explorer $G_{n,T}$ en au plus $2n - T - 1$ unités de temps.

Preuve Soit $n \geq 3$, $1 \leq T \leq \lfloor \frac{n-1}{2} \rfloor$ et $G_{n,T}$ un graphe dynamique T -intervalle-connexe qui a comme graphe sous-jacent A_n . Sans perte de généralité, supposons que l'agent A est sur le sommet s_0 au temps $t = 0$. Soit A_G un agent qui depuis le point de départ exécute EXPLORE-ANNEAU (*gauche*) et A_D un agent qui depuis le point de départ exécute EXPLORE-ANNEAU (*droite*). Supposons qu'il existe une arête où l'adversaire bloque A_G et A_D en même temps pendant un temps $t' \geq 0$, soit e cette arête. Soit u le sommet qui est à droite de e et v le sommet à gauche de e . Soit $A_{G'}$ un agent qui depuis le point de départ exécute EXPLORE-ANNEAU (*gauche*) et visite le maximum de sommet vers la gauche avant d'exécuter EXPLORE-ANNEAU (*droite*) pour arriver au plus tard sur le sommet u avant que A_D traverse e . Soit $A_{D'}$ un agent qui depuis le point de départ exécute EXPLORE-ANNEAU (*droite*) et visite le maximum de sommet vers la droite avant d'exécuter EXPLORE-ANNEAU (*gauche*) pour arriver au plus tard sur le sommet v avant que A_D traverse e . Soit $A_{G''}$ un agent qui depuis le point de départ exécute EXPLORE-ANNEAU (*gauche*) et fait une unité de temps de plus que $A_{G'}$ vers la gauche avant d'exécuter EXPLORE-ANNEAU (*droite*) et $A_{D''}$ un agent qui depuis le point de départ exécute EXPLORE-ANNEAU (*droite*) et fait une unité de temps de plus que $A_{D'}$ vers la gauche avant d'exécuter EXPLORE-ANNEAU (*gauche*). Comme le graphe est toujours connexe, alors $A_{G'}$ ou $A_{D'}$ réussira à explorer le graphe dynamique en un temps fini. Soit \mathcal{T} le temps d'exploration du plus rapide entre $A_{G'}$ et $A_{D'}$.

1) Si aucune arête n'est supprimée entre le l'intervalle de temps $[0, \mathcal{T}[$ alors e est quelconque et $t' = 0$. Donc A_G et A_D réussiront à explorer le graphe en $n - 1$ unités de temps.

2) Si une arête est supprimée dans le graphe dans l'intervalle de temps $[0, \mathcal{T}[$, soit e' cette arête. Soit B un agent qui exécute l'algorithme 5 avec la direction du premier agent qui arrive sur e' . Supposons que B fait demi-tour une fois sur le sommet avant l'arête e' . Soit B n'est jamais bloqué après qu'il a fait demi-tour jusqu'au temps \mathcal{T} et qu'il explore tout le graphe en au plus $\lfloor \frac{3n}{2} \rfloor$ unités de temps. Soit il est bloqué après \mathcal{T} et donc $\mathcal{T} \leq \lfloor \frac{3n}{2} \rfloor$.

3) Si au moins deux arêtes sont supprimées dans le graphe dans l'intervalle de temps $[0, \mathcal{T}[$. Soit e l'arête bloquante de A_G et A_D et t_1 le temps où le premier entre A_G et A_D arrive sur e . Soit d_j^I la distance parcourut par l'agent A_j et α_j^I le nombre de fois qu'il est bloqué sur un sommet durant l'intervalle de temps I . Soit $\alpha_{j,k}^I$ le nombre de fois où l'adversaire bloque A_j et A_k en même temps durant l'intervalle de temps I . Soit β^I le nombre de fois où toutes les arêtes du graphe sous-jacent sont présentes durant l'intervalle de temps I .

a) Si e est supprimée au moins de t_1 à $t_1 + n - 1$ alors $\mathcal{T} \leq t_1 + n - 1$ (1).

$$\text{Soit } I_1 = [0, t_1[\text{ on a. } t_1 = \begin{cases} d_G^{I_1} + \alpha_G^{I_1} & (2) \\ d_D^{I_1} + \alpha_D^{I_1} & (3) \end{cases}$$

$$\alpha_G^{I_1} + \alpha_D^{I_1} + \beta^{I_1} \leq t_1 \quad (4)$$

$$\text{Or on a } d_G^{I_1} + d_D^{I_1} \leq n - 1 \quad (5) \quad \text{et cas 3)} \rightarrow \beta^{I_1} \geq T - 1 \quad (6)$$

$$(1)+(2)+(3)+(4)+(5)+(6) \rightarrow \mathcal{T} \leq 2n - T - 1$$

b) Si e est supprimée pendant un temps $< n - 1$. Soit t_2 le temps que met l'arête e après t_1 pour réapparaître. Soit $I_1 = [0, t_1[$, $I_2 = [t_1, t_1 + t_2[$, $I_{1,2} = [0, t_1 + t_2[$, $I_3 = [t_1 + t_2, T[$ et $I = [0, T[$.

Sur I_1 on a

$$t_1 \geq \alpha_{G''}^{I_1} + \alpha_{D''}^{I_1} - \alpha_{G'',D''}^{I_1} + \beta^{I_1} \quad (1)$$

$$t_1 \geq \alpha_G^{I_1} + \alpha_D^{I_1} + \alpha_{G',D'}^{I_1} + \beta^{I_1} \quad (2)$$

$$t_1 = \begin{cases} d_G^{I_1} + \alpha_G^{I_1} & (3) \\ d_D^{I_1} + \alpha_D^{I_1} & (4) \end{cases}$$

$$(1) + (2) + (3) + (4) \rightarrow \alpha_{D''}^{I_1} + \alpha_{G''}^{I_1} + 2\beta^{I_1} \leq d_G^{I_1} + d_D^{I_1} \quad (5)$$

Sur $I_{1,2}$ on a

$$t_1 + t_2 = \begin{cases} d_{G''}^{I_{1,2}} + \alpha_{G''}^{I_{1,2}} & (6) \\ d_{D''}^{I_{1,2}} + \alpha_{D''}^{I_{1,2}} & (7) \end{cases}$$

$$\text{Remarquons que par définition, } d_{G''}^{I_{1,2}} \leq d_{G'}^{I_{1,2}} + 1 \quad (8) \quad \text{et } d_{D''}^{I_{1,2}} \leq d_{D'}^{I_{1,2}} + 1 \quad (9)$$

$$(6) + (7) + (8) + (9) \rightarrow 2(t_1 + t_2) \leq d_{G'}^{I_{1,2}} + d_{D'}^{I_{1,2}} + \alpha_{D'}^{I_{1,2}} + \alpha_{G'}^{I_{1,2}} + 2 \quad (10)$$

Remarquons que sur $[t_1, t_1 + t_2[$, $A_{G''}$ et $A_{D''}$ ne sont pas bloqués car e est absente, donc

$$\alpha_{G''}^{I_{1,2}} = \alpha_{G'}^{I_1} \quad (11) \quad \text{et} \quad \alpha_{D''}^{I_{1,2}} = \alpha_{D'}^{I_1} \quad (12)$$

$$(5) + (10) + (11) + (12) \rightarrow 2(t_1 + t_2) + 2\beta^{I_1} \leq d_G^{I_1} + d_D^{I_1} + d_G^{I_{1,2}} + d_D^{I_{1,2}} + 2 \quad (13)$$

Sur I_3 on a

$$\mathcal{T} - (t_1 + t_2) \geq \alpha_{G'}^{I_3} + \alpha_{D'}^{I_3} + \beta^{I_3} \quad (14)$$

$$\mathcal{T} - (t_1 + t_2) = \begin{cases} d_{G'}^{I_3} + \alpha_{G'}^{I_3} & (15) \\ d_{D'}^{I_3} + \alpha_{D'}^{I_3} & (16) \end{cases}$$

$$(14) + (15) + (16) \rightarrow \mathcal{T} - (t_1 + t_2) + \beta^{I_3} \leq d_{G'}^{I_3} + d_{D'}^{I_3} \quad (17)$$

$$(17) + \frac{1}{2}(13) \rightarrow \mathcal{T} + \beta^{I_1} + \beta^{I_3} \leq \frac{1}{2}(d_G^{I_1} + d_D^{I_1} + d_{G'}^{I_{1,2}} + d_{D'}^{I_{1,2}}) + d_{G'}^{I_3} + d_{D'}^{I_3} + 1 \quad (18)$$

Notons que $\beta^{I_1} + \beta^{I_3} = \beta^I$

$$d_{D'}^{I_{1,2}} = 2x + d_G^{I_1} \quad (19), \text{ avec } x \text{ la distance parcourue par } A_{D'} \text{ avant de faire demi-tour}$$

$$d_{G'}^{I_{1,2}} = 2y + d_D^{I_1} \quad (20), \text{ avec } y \text{ la distance parcourue par } A_{G'} \text{ avant de faire demi-tour}$$

$$d_{D'}^{I_3} = d_D^{I_1} - x \quad (21) \quad d_{G'}^{I_3} = d_G^{I_1} - y \quad (22)$$

$$d_G^{I_1} + d_D^{I_1} \leq n - 1 \quad (23) \quad \text{et cas 3)} \rightarrow \beta^I \geq T - 1 \quad (24)$$

$$(18) + (19) + (20) + (21) + (22) + (23) + (24) \rightarrow \mathcal{T} \leq 2n - T - 1$$

□

5.2. Le cas où l'agent ne connaît pas la dynamique du graphe

Dans cette section nous allons donner une borne supérieure sur le temps d'exploration des graphes dynamiques T-intervalle-connexes δ -récurrents qui ont comme graphe sous-jacent A_n . Pour cela, nous allons supposer que l'agent connaît le graphe sous-jacent mais ne connaît pas la dynamique du graphe.

Théorème 4 *L'algorithme EXPLORE-ANNEAU permet d'explorer tous les graphes dynamiques T-intervalle-connexes δ -récurrents qui ont A_n comme graphe sous-jacent en au plus $\lceil \frac{n-1}{\max\{1, T-1\}} \rceil (\delta - 1) + n - 1$ unités de temps.*

Preuve Comme le graphe dynamique est T-intervalle-connexe donc tout agent exécutant l'algorithme 5 ne peut être bloqué qu'au plus une fois toutes les $\max\{1, T - 1\}$ unités de temps. Le graphe dynamique est aussi δ -récurrent donc chaque arête du graphe sous-jacent va apparaître au moins une fois toutes les δ unités de temps. Ce qui veut dire que tout agent exécutant l'algorithme 5 peut être appelé à attendre pendant $\delta - 1$ unités de temps sur un sommet qu'au plus une fois toutes les $\max\{1, T - 1\}$ unités de temps.

Wade

Comme le graphe dynamique est de n sommets, un agent exécutant l'algorithme 5 va payer au plus $\lceil \frac{n-1}{\max\{1, T-1\}} \rceil (\delta - 1)$ unités de temps d'attente et $n - 1$ unités de temps d'exploration. Donc l'agent va payer au plus $\lceil \frac{n-1}{\max\{1, T-1\}} \rceil (\delta - 1) + n - 1$ unités de temps pour explorer le graphe. \square

6. Conclusion

Dans cet article, nous avons étudié le problème d'exploration des graphes dynamiques T-intervalle-connexes qui ont comme graphe sous-jacent un anneau de taille n , en considérant le cas où l'agent connaît la dynamique du graphe et le cas où il ne le connaît pas.

La suite de ces travaux serait de continuer l'étude des graphes dynamiques T-intervalle-connexes en étendant les graphes sous-jacents considérés à une famille plus large que la famille des anneaux.

7. Bibliographie

- [1] A. CASTEIGTS, P. FLOCCINI, W. QUATTROCIOCHI, N. SANTORO, « Time-varying graphs and dynamic networks », *CoRR*, n° 1012.0009, 2010.
- [2] C. COOPER, A. M. FRIEZE, « Crawling on simple models of web graphs », *Internet Mathematics*, vol. 1(1), 2003.
- [3] A. DESSMARK, A. PELC, « Optimal graph exploration without good maps », *Theor. Comput. Sci.*, vol. 326(1-3), p. 343–362, 2004.
- [4] A. Ferreira. Building a Reference Combinatorial Model for Dynamic Networks :Initial Results in Evolving Graphs. INRIA, RR-5041 (2003)
- [5] P. Flocchini, M. Kellett, P. C. Mason, and N. Santoro. Searching for black holes in subways. *Theory of Computing Systems*, 50(1), pages 158- ?184, 2012.
- [6] P. FLOCCINI, B. MANS, N. SANTORO, « Exploration of periodically varying graphs », *20th Intl Symposium on Algorithms and Computation (ISAAC)*, vol. LNCS 5878, p. 534–543, 2009.
- [7] D. ILCINKAS, A.M. WADE, « On the Power of Waiting When Exploring Public Transportation Systems », *16th International Conference On Principles Of Distributed Systems (OPODIS)*, vol. LNCS 7109, p. 451–464, 2011.
- [8] T. Izumi, Y. Yamauchi, and S. Kamei. Brief Announcement : Mobile Agent Rendezvous on Edge Evolving Rings. In *14th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS)*, LNCS 7596, pages 92–94, 2012.
- [9] F. KUHN, N.A. LYNCH, R. OSHMAN, « Distributed computation in dynamic networks », *STOC*, vol. STOC, p. 513–522, 2010.
- [10] P. PANAITÉ, A. PELC, « Exploring Unknown Undirected Graphs », *J. Algorithms*, vol. 33(2), p. 281–295, 1999.
- [11] O. REINGOLD, « Undirected st-connectivity in log-space », *37th ACM Symposium on Theory of Computing (STOC)*, p. 376–385, 2005.
- [12] C. E. SHANNON, « Presentation of a maze-solving machine », *8th Conf. of the Josiah Macy Jr. Found. (Cybernetics)*, p. 173–180, 1951.

CNRIA'2013

1. Introduction

Dans les réseaux *ad hoc*, la principale source de consommation de ressources sont les émissions et réceptions de messages. La solution de communication la plus utilisée est la diffusion. C'est une technique simple qui nécessite peu de calcul. Mais elle est coûteuse et peut entraîner une saturation du réseau. Pour optimiser les communications, une solution est de structurer le réseau en *clusters* [4]. En outre, les réseaux *ad hoc* peuvent être également dynamiques. Un nœud peut entrer ou sortir du réseau à tout moment et peut donc induire des fautes transitoires dans le réseau. Donc, concevoir des applications tolérantes aux fautes transitoires est plus que nécessaire pour des applications de réseaux *ad hoc* souvent critiques. L'auto-stabilisation, introduit par *Dijkstra* en 1974, est une propriété intéressante de tolérance aux pannes transitoires [6]. Un système est dit auto-stabilisant si quelle que soit sa configuration de départ, il est garanti d'arriver à une configuration légale en un nombre fini d'étapes. Suite à l'apparition d'une faute transitoire, le système de lui-même, retrouve une configuration légale.

Il existe plusieurs solutions de *clustering* auto-stabilisantes. Elles se basent soit sur un modèle à états [10, 11, 5, 3] soit sur un modèle à passage de messages [12, 9, 13]. Les solutions basées sur un modèle à états bien qu'elles offrent l'avantage d'éviter l'envoi et la réception de messages, la perte de messages, elles ne sont pas réalistes dans le cadre des réseaux *ad hoc*. Les solutions basées sur du passage de messages, plus réalistes, sont souvent coûteuses en termes d'échange de messages.

Dans cet article, nous proposons une implémentation de notre proposition décrite dans [1]. Notre solution est auto-stabilisante et ne nécessite aucune initialisation. Elle se base sur l'information provenant des nœuds voisins à distance 1 pour construire des *clusters* de diamètre au plus $2k$. Partant d'une configuration quelconque, le réseau converge à un état stable au plus en $n + 2$ transitions et nécessite une occupation mémoire d'au plus $n * \log(2n + k + 3)$. Nous donnons une comparaison analytique et expérimentale entre notre approche et les solutions de *clustering* auto-stabilisantes à k sauts [5, 3, 13].

La suite de l'article est organisée comme suit. Dans la section 2, nous étudions quelques solutions de *clustering* auto-stabilisantes. La section 3 présente notre contribution. A la section 4, nous rappelons le principe d'exécution de notre approche. Dans la section 5, nous comparons analytiquement notre approche avec les solutions basées sur un modèle à états. Une implémentation de notre approche sous le simulateur OMNeT++ est proposée dans la section 6 de même qu'une comparaison avec une solution bien connue basée sur un modèle à passage de messages. Une conclusion et des perspectives sont données dans la section 7.

2. État de l'art

Plusieurs propositions de *clustering* auto-stabilisantes ont été faites dans la littérature [5, 3, 9, 11, 10, 13, 12].

Les approches auto-stabilisantes [9, 11, 10, 12] construisent des *clusters* à 1 saut. Mitton *et al.* [12] utilisent comme métrique la *densité* qu'ils définissent comme étant le nombre de connexions entre voisins sur le nombre de voisins afin de minimiser la reconstruction de la structure en cas de faible changement de topologie. Chaque nœud

calcule sa densité et la diffuse à ses voisins. Le nœud avec la plus grande densité devient *clusterhead*. Flauzac et al. [9], ont proposé un algorithme auto-stabilisant qui combine la découverte de topologie et de *clustering* grâce à un seul type de message échangé entre voisins. Un nœud devient *clusterhead* s'il possède la plus grande identité parmi tous ses voisins. Johnen et al. [10] ont proposé un algorithme auto-stabilisant qui construit des *clusters* de taille fixe. Ils attribuent un poids à chaque nœud et fixent un paramètre *SizeBound* qui représente le nombre maximal de nœuds dans un *cluster*. Un nœud ayant le poids le plus élevé devient *clusterhead* et collecte dans son *cluster* jusqu'à *SizeBound* nœuds. Dans [11], Johnen et al. ont étendu leur proposition décrite dans [10] pour apporter la notion de *robustesse* qui est la propriété qui assure que partant d'une configuration quelconque, le réseau est partitionné après un *round* asynchrone. Et durant la phase de convergence, il reste toujours partitionner et vérifie un prédicat de sureté.

Les approches auto-stabilisantes [5, 3, 13] construisent des *clusters* à k sauts.

Le critère d'identité minimale est utilisé dans [5] pour construire des *clusters* à k sauts. Étant donné un graph G , les auteurs construisent un sous-ensemble D , appelé *k-dominating*, où tout nœud n'appartenant pas à D est au plus à distance k d'au moins d'un membre de D . Ils considèrent D comme étant l'ensemble des *clusterhead* et partitionnent le graphe en *clusters* de diamètre au plus $2k$. Cet algorithme se stabilise en $O(n)$ *rounds* et $O(n^2)$ étapes et nécessite un espace mémoire de $\log(n)$ où n est le nombre de nœuds du réseau. Datta et al. [3], utilisant une métrique arbitraire, ont proposé un algorithme auto-stabilisant basé sur un modèle à états nommé *k-Clustering*. Cet algorithme est très lent, il s'exécute en $O(n * k)$ *rounds* et nécessite $O(\log(n) + \log(k))$ espace mémoire par processus, où n est le nombre de nœuds du réseau. Dans [13], Miton et al. étendent leurs travaux décrits dans [12] pour proposer un algorithme robuste de *clustering* auto-stabilisant à k sauts. Chaque nœud recueille les informations dans son voisinage à distance $k + 1$, calcule sa *k-densité* et la diffuse à son voisinage à distance k . Le nœud avec la plus grande densité devient *clusterhead*. Avec cet algorithme, si un nœud est trop mobile, il risque de ne se rattacher à aucun *cluster*.

Les solutions basées sur un modèle à états [10, 11, 5, 3] ne sont pas réalistes pour le cadre des réseaux ad hoc. Par contre, les solutions basées sur du passage de messages [12, 13] sont coûteuses en termes d'échange de messages.

3. Contribution

Notre solution [1] est basée sur un modèle à passage de messages. Elle est asynchrone, distribuée et auto-stabilisante. Elle ne nécessite aucune initialisation et gère la corruption mémoire. Nous utilisons le critère d'identité maximale pour le choix du *clusterhead*. Notre algorithme structure le réseau en *clusters* disjoints deux à deux de diamètre au plus égal à $2k$. Elle combine la découverte de voisinage et le *clustering* en une seule phase. Elle se base uniquement sur l'information provenant des nœuds voisins situés à distance 1 par le biais d'échange périodique de messages. Notre approche se stabilise au plus en $n + 2$ transitions et nécessite une occupation mémoire d'au plus $n * \log(2n + k + 3)$. Nous comparons analytiquement notre approche à celles basées sur un modèle à états [5, 3]. Par simulations, nous la comparons avec la solution décrite dans [13] basée sur un modèle à passage de messages.

4. Principe d'exécution de notre approche

Dans cette section, nous décrivons le principe d'exécution de notre approche. Les principales définitions et la description formelle de notre algorithme sont données dans [1].

La solution que nous proposons se déroule comme suit :

Dès la réception d'un message *hello*, un nœud u exécute l'algorithme de clustering décrit dans [1]. Dans cet algorithme, u exécute trois étapes consécutives. D'abord il effectue la mise à jour de sa table de voisinage, ensuite la vérification de la cohérence de ses variables locales pour gérer la corruption mémoire et en dernier la phase de *clustering*. A la fin de cette exécution, u envoie un message *hello* à tous ses voisins.

Durant l'étape de *clustering*, chaque nœud compare son identité à celle de tous ses voisins. Un nœud u s'élit *clusterhead* avec un statut de *CH* (*ClusterHead*) s'il la plus grande identité parmi tous les nœuds de son *cluster*. Si le nœud u découvre un voisin v avec une plus grande identité que la sienne, alors il devient membre du même *cluster* que v avec un statut de *SN* (*Simple Node*). Si un nœud reçoit un message provenant d'un voisin appartenant à un autre *cluster*, alors il devient un nœud de passage avec un statut de *GN* (*Gateway Node*). Comme les messages *hello* contiennent la distance entre chaque nœud u et son *clusterhead*, alors u peut déterminer si le diamètre maximal du *cluster* est atteint. Ainsi, u pourra choisir un autre *cluster* si les k sauts sont atteints.

5. Étude analytique

Dans [2], nous avons prouvé que partant d'une configuration quelconque, une configuration légale est atteinte au plus en $n + 2$ transitions. De plus, notre approche nécessite une occupation mémoire de $\log(2n + k + 3)$ par nœud ; donc $n * \log(2n + k + 3)$ espace mémoire au plus pour un réseau de n nœuds.

Le tableau 1 illustre une comparaison du temps de stabilisation, de l'occupation mémoire et du voisinage (distance d'écoute ou de lecture) entre notre approche et celles basées sur un modèle à états.

	Temps de stabilisation	Occupation mémoire / nœud	Voisinage
Notre approche	$n + 2$	$\log(2n + k + 3)$	1 saut
Datta et al. [5]	$O(n), O(n^2)$	$O(\log(n))$	$k \text{ sauts}$
Datta et al. [3]	$O(n * k)$	$O(\log(n) + \log(k))$	$k+1 \text{ sauts}$

Tableau 1 – Comparaison entre temps de stabilisation et occupation mémoire

Notre temps de stabilisation ne dépend pas du paramètre k contrairement à la solution [3]. Comme nous avons une unique phase de découverte de voisinage et ce construction de *clusters* à k sauts, nous avons un unique temps de stabilisation contrairement à l'approche décrite dans [5]. De plus nous considérons uniquement un voisinage à distance 1 à l'opposé des solutions [5, 3].

6. Étude expérimentale

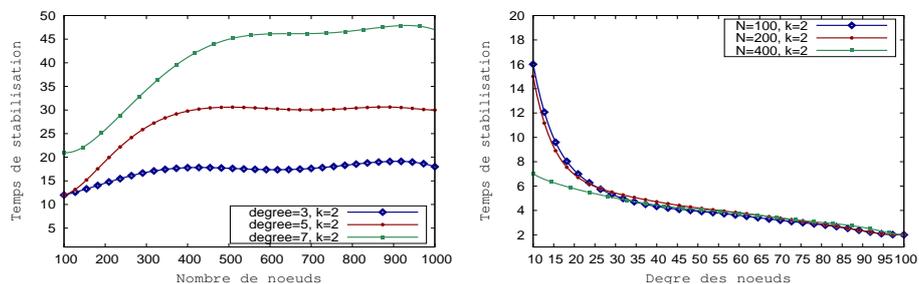
Dans cette section, nous donnons une implémentation et une évaluation de notre approche par simulation. De plus, nous comparons notre approche avec la proposition de Mitton *et al.* [13].

Comme nous l'avons montré dans [2], notre réseau converge au plus en $n + 2$ transitions. Ceci correspond au cas le plus défavorable dans une topologie où les nœuds forment une chaîne ordonnée. Or, un réseau *ad hoc* est caractérisé par une topologie dense et aléatoire. C'est ainsi que nous faisons recours à une campagne de simulations pour évaluer les performances moyennes de notre algorithme.

Nous utilisons le simulateur *OMNeT++* [14] et la librairie *SNAP* [8] comme générateur de graphes. Toutes nos simulations sont effectuées sous *Grid'5000* [7].

6.1. Impact de la densité de connexité du réseau sur le temps de stabilisation

Ici, nous étudions l'impact du degré (nombre de voisins) et du nombre de nœuds sur le temps de stabilisation du réseau. Nous fixons un degré de 3, 5 et 7 et nous faisons varier le nombre de nœuds de 100 à 1000 par pas de 100. Avec la librairie *SNAP*, nous générons des graphes *d-réguliers*, où d est le degré des nœuds. Nous fixons $k = 2$ et pour chaque pas, nous effectuons plusieurs séries de simulations. Nous fournissons des valeurs moyennes. Au niveau de la figure 1(a), nous remarquons que le temps de stabilisation augmente légèrement avec l'augmentation du nombre de nœuds et varie peu à partir d'une certaine taille du réseau. De plus, nous notons que pour des topologies quelconques, le temps de stabilisation moyen est très en dessous de $n + 2$, valeur formelle prouvée dans le pire des cas.



(a) Temps de stabilisation en fonction du nombre de nœuds (b) Temps de stabilisation en fonction du degré des nœuds

Figure 1 – Impact de la densité et du nombre de nœuds sur le temps de stabilisation

Pour mieux observer l'impact de la densité du réseau comme illustré avec la figure 1(b), nous fixons le nombre de nœuds et nous faisons varier le degré. Nous observons que pour un nombre de nœuds fixé à 100, 200 et 400, plus le degré des nœuds augmente, plus le temps de stabilisation diminue. En effet, si le degré augmente, les nœuds ayant les plus grandes identités ont plus de voisins. Donc, ils absorbent plus de nœuds dans leurs *clusters* durant chaque transition. Ainsi, avec notre approche, plus le réseau est dense, plus nous avons de meilleurs temps de stabilisation. Or, les réseaux *ad hoc* sont souvent caractérisés par une forte densité.

6.2. Étude du passage à l'échelle

Pour étudier le passage à l'échelle de notre approche, nous faisons varier le nombre de nœuds dans le réseau en même temps que la densité de connexité. Pour $k = 2$, nous fai-

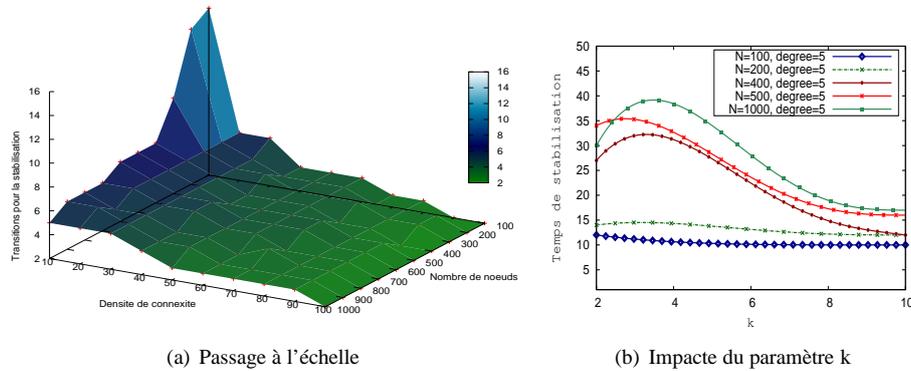


Figure 2 – Passage à l'échelle - Impact du paramètre k

sons évoluer le nombre de nœuds de 100 à 1000 par pas de 100. Et pour chaque valeur de nombre de nœuds, nous faisons varier la densité du réseau de 10% à 100% par pas de 10. Nous obtenons la courbe 3D de la figure 2(a). Nous remarquons que sauf pour de faibles densités (10% et 20%), le temps de stabilisation varie légèrement avec l'augmentation du nombre de nœuds. Et cas de faibles densités, nous observons un pic. Mais avec l'augmentation du nombre de nœuds, le temps de stabilisation diminue et nous observons le même phénomène que la Figure 1(b). Avec cette série de simulations, nous pouvons soulever deux remarques. (i) Seule la densité de connectivité est le facteur déterminant avec notre approche. (ii) En moyenne, pour des réseaux avec une topologie quelconque, le temps de stabilisation est très inférieur à celui du pire des cas ($n + 2$ transitions). Rappelons que ce pire cas est une topologie où les nœuds forment une chaîne ordonnée comme nous l'avons prouvé dans [2].

6.3. Impact du paramètre k

Pour analyser l'impact du paramètre k , nous fixons arbitrairement un degré de 5 et nous considérons des réseaux de taille 100, 200, 400, 500 et 1000 nœuds. Pour chaque taille de réseau, nous faisons varier la valeur de k de 2 à 10. La figure 2(b) montre le temps de stabilisation en fonction de la valeur de k . Nous observons une diminution du temps de stabilisation avec l'augmentation de la valeur de k . En effet, comme les messages *hello* échangés contiennent la valeur de la distance k , si k augmente, le champs d'influence des nœuds avec une plus grande identité augmente. Les nœuds effectuent moins de transitions pour se *fixer* à un *CH*. Avec de petites valeurs du paramètre k , nous avons des *clusters* de faibles diamètres. Donc, il nécessite plus de transitions pour atteindre un état stable dans tous les *clusters*. Notons que quelle que soit la valeur du paramètre k et pour des graphes de réseaux totalement aléatoires, nous obtenons des temps de stabilisation très inférieurs au pire des cas ($n + 2$ transitions).

6.4. Comparaison avec une approche basée sur un modèle à passage de messages

Il existe plusieurs propositions basées sur un modèle à passage de messages. Nous comparons notre approche avec une des solutions bien connue basée sur un modèle à passage de messages [13]. La comparaison s'effectue en termes de nombre de messages et

de nombre de *clusters*. De même que nous approchons, nous implémentons l'approche [13] dans le simulateur *OMNeT++*. Toutes les simulations sont effectuées en considérant les mêmes graphes de réseaux.

Pour évaluer le nombre de messages échangés, nous fixons un degré de 3 et 6 et nous faisons varier la taille du réseau de 100 à 1000 nœuds. Nous fixons $k = 2$ et pour chaque taille de réseau, nous déroulons une série de plusieurs simulations. Nous donnons le nombre moyen de messages échangés pour atteindre un état légal (formation des *clusters* stables). Comme illustré au niveau de la figure 3(a), notre approche génère beaucoup moins de messages que [13]. En effet, nous réduisons d'au moins 50% le nombre de messages envoyés par rapport à [13]. La principale raison est que notre approche se base uniquement sur l'information provenant du voisinage à distance 1 pour construire des *clusters* de diamètre au plus $2k$. Alors que pour la solution [13], chaque nœud doit connaître le voisinage à distance $k + 1$, calcule sa *k-densité* et la diffuse dans le voisinage à distance k . Ceci est très coûteux en terme d'échange de messages.

Nous avons aussi évalué le nombre de *clusters* obtenus avec les deux approches. Comme illustré au niveau de la figure 3(b), nous construisons plus de *clusters* en générant moins de messages. En outre nos *clusters* sont moins denses. Donc ils sont faciles à gérer par le *clusterhead* qui consomme ainsi moins de ressources. Des *clusters* moins denses entraînent moins de communications et donc une optimisation de la consommation de ressources.

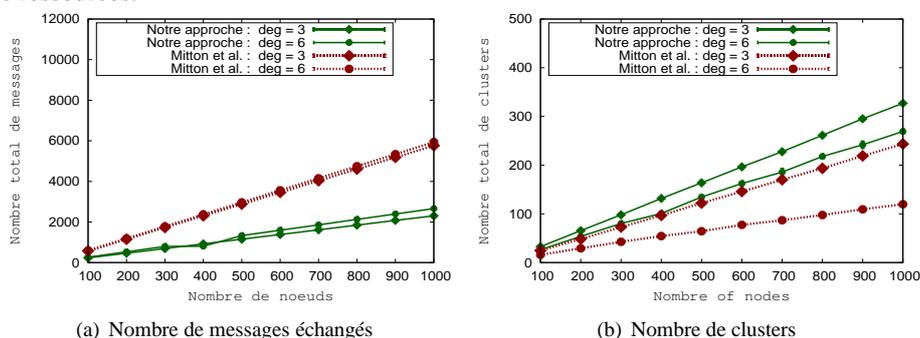


Figure 3 – Comparaison avec Mitton et al. [13]

Notons que nous avons effectué les comparaisons pour $k = 2$. Quand la valeur de k augmente, le champ de diffusion de la solution [13] augmente. Or, notre approche se base toujours sur l'information du voisinage à distance 1 pour construire des *clusters* à k sauts.

7. Conclusion

Dans cet article, nous avons proposé une implémentation de notre approche *clustering* ainsi qu'une étude comparative analytique et expérimentale entre solutions de *clustering* auto-stabilisantes. Sans initialisation, notre solution n'utilise que des informations provenant des nœuds voisins. Elle combine la découverte du voisinage et la construction des *clusters*. Nous avons montré que l'état stable est atteint en au plus $n + 2$ transitions et nécessite au plus une occupation mémoire de $n * \log(2n + k + 3)$. Nous avons donné une

comparaison analytique et expérimentale de notre approche par rapport aux solutions existantes. Par simulations, nous avons évalué les performances de notre proposition. Notre solution génère moins de messages et présente de meilleurs temps de stabilisation.

Dans nos futurs travaux, nous comptons faire une implémentation et une étude de cas de notre proposition dans les plateformes de réseaux de capteurs *CapSec* et *SensLab*. De plus, nous comptons proposer une approche de routage basée sur la structure de nos *clusters*.

Remerciements

Ces travaux ont été effectués dans le cadre du projet CPER CapSec ROFICA co-financé par la région de Champagne-Ardenne et le FEDER. Nous remercions également le Centre de Calcul de Champagne-Ardenne ROMEO et Grid'5000.

8. Bibliographie

- [1] M. Ba, O. Flauzac, B. S. Hagggar, F. Nolot, and I. Niang. Clustering auto-stabilisant à k sauts dans les réseaux ad hoc. In *Actes du 4e CNRIA*, 2012. 8p.
- [2] M. Ba, O. Flauzac, B. S. Hagggar, F. Nolot, and I. Niang. Self-stabilizing k -hops clustering algorithm for wireless ad hoc networks. In *7th ACM IMCOM*, pages 38 :1–38 :10, 2013.
- [3] E. Caron, A. Datta, B. Depardon, and L. Larmore. A self-stabilizing k -clustering algorithm for weighted graphs. *J. Parallel Distrib. Comput.*, pages 1159–1173, 2010.
- [4] A. Datta, L. Larmore, and P. Vemula. Self-stabilizing leader election in optimal space. In *Stabilization, Safety, and Security of Distributed Systems*, pages 109–123. 2008.
- [5] Ajoy K. Datta, Stéphane Devismes, and Lawrence L. Larmore. A self-stabilizing $o(n)$ -round k -clustering algorithm. In *28th IEEE SRDS*, pages 147–155, 2009.
- [6] E. W Dijkstra. Self-stabilizing systems in spite of distributed control. *Commun. ACM*, pages 643–644, 1974.
- [7] Cappello F. *et al.* Grid'5000 : A large scale and highly reconfigurable grid experimental testbed. In *GRID*, pages 99–106, 2005.
- [8] Leskovec J. *et al.* SNAP : Stanford Network Analysis Platform, 2007. <http://snap.stanford.edu/index.html>.
- [9] O. Flauzac, B. S. Hagggar, and F. Nolot. Self-stabilizing clustering algorithm for ad hoc networks. *ICWMC*, pages 24–29, 2009.
- [10] C. Johnen and L. H. Nguyen. Self-stabilizing construction of bounded size clusters. *ISPA*, pages 43–50, 2008.
- [11] C. Johnen and L. H. Nguyen. Robust self-stabilizing weight-based clustering algorithm. *Theoretical Computer Science*, pages 581 – 594, 2009.
- [12] N. Mitton, A. Busson, and E. Fleury. Self-organization in large scale ad hoc networks. 2004.
- [13] N. Mitton, E. Fleury, I. Guerin L., and S. Tixeuil. Self-stabilization in self-organized multihop wireless networks. In *ICDCSW*, pages 909–915, 2005.
- [14] András Varga and Rudolf Hornig. An overview of the OMNeT++ simulation environment. In *Simutools*, pages 60 :1–60 :10, 2008.

Actes du 5^{eme} Colloque National sur la Recherche en Informatique et ses Applications - Ziguinchor, Sénégal, Avril 2013

Amélioration de l'Algorithme P.E.G

par Minimisation des Trapping-Set TS(5,3) sur les Codes LDPC

Madiagne DIOUF^{1,2} — David DECLERCQ² — Samuel OUYA¹

¹ Laboratoire Informatique Réseaux & Télécom (LIRT)
Ecole Supérieure Polytechnique / Université Cheikh Anta DIOP

² Equipes Traitement de l'Information et Systèmes (ETIS)
ENSEA / Université de Cergy-Pontoise / CNRS UMR 8051
6,avenue du Ponceau, 95000 Cergy-Pontoise, France

{ madiagne.diouf,declercq}@ensea.fr, samuel.ouya@gmail.com

RÉSUMÉ. L'algorithme PEG (Progressive Edge-Growth est très utilisé pour la minimisation du nombre de cycle lors de la construction de matrices dans les codes LDPC. Dans cet article nous proposons une amélioration de l'algorithme de construction P.E.G (Progressive Edge Growth) par réduction des trapping-set TS(5,3) pour des matrices de plus faible cycle $G \geq 6$. Nous montrons par des simulations, avec des rendements et tailles code différentes, que cette modification améliore les performances des codes LDPC pour un décodeur basé sur la propagation de croyance.

ABSTRACT. The algorithm PEG (Progressive Edge-Growth is widely used for the minimization of the number of cycles in the construction of matrices in LDPC codes. In this paper we propose an improved algorithm for constructing PEG (Progressive Edge Growth) by reducing a trapping-set TS(5,3) for the matrices by $G \geq 6$. We show by simulations with different yields and size code, this modified algorithm improves performance of LDPC codes with the belief propagation decoder

MOTS-CLÉS : LDPC, Cycle, Propagation de Croyance, croissance progressive bord, Graphe de Tanner

KEYWORDS : LDPC, Cycle, Belief-Propagation, Progressive Edge Growth(PEG) Trapping-Set, Tanner Graph

1. INTRODUCTION

Les codes Low Density Parity Check(LDPC) sont inventés par Gallager en 1963[1], mais ils furent oubliés longtemps puis repris par Mackay and Neal[2] qui améliore le décodeur par l'introduction d'itération. Richardson et Urbanke montrent que les performances des systèmes de codes LDPC s'approchent désormais de la limite de Shannon [3], dès lors que la taille des mots de codes est très grande. Par contre, dès que la taille des mots de codes devient plus faible, ce qui est le cas dans pratiquement toutes les applications pratiques, ces performances se dégradent. Plusieurs techniques de conception de codes LDPC ont été proposées, dont l'objectif est d'avoir de bonnes performances avec l'algorithme de décodage itératif "Belief-Propagation". Il a été montré en particulier que les meilleures matrices de contrôle qui définissent le code LDPC doivent comporter peu ou pas de cycles courts[1][2], ni certaines de leurs combinaisons appelées trapping sets[4]. Plusieurs algorithmes ont été proposés dans le but d'augmenter la taille des cycles et de minimiser leur nombre, dont le plus connu est l'algorithme PEG (Progressive Edge-Growth [5]), qui permet d'obtenir des girth $g \geq 6$, même pour de très petites tailles de codes. Cet algorithme ne traitant que les cycles n'est pas suffisant pour atteindre les meilleures performances, surtout dans les très faibles taux d'erreur, il faut aussi éviter certaines combinaisons.

Dans cet article nous proposons une amélioration de l'algorithme PEG, par une réduction des trapping TS(5,3), afin de construire de meilleurs codes LDPC à taille finie. Dans la section 2, nous rappelons l'algorithme PEG original avec ces définitions et notations qui seront utilisées. Sur la section 3 nous décrivons le principe de réduction des trapping TS(5,3) pour un code de degrés variable $d_v = 3$. Nous terminerons par une comparaison des résultats des simulations.

2. ALGORITHME PEG ORIGINAL

Un code LDPC dont les mots codes sont de taille N est défini par sa matrice de parité H . Nous appelons degré d'un noeud variable v_j , noté d_{v_j} le poids de la j^{eme} colonne de la matrice de parité, le poids de la i^{eme} ligne est le degré du noeud de parité c_i . Nous utiliserons les mêmes notations et définitions dans [5]-[7]. Pour une variable v_j , $\mathcal{N}_{v_j}^l$ représente l'ensemble des noeuds de parité rencontrés sur l'arbre de calcul du sommet jusqu'à la profondeur l et $\overline{\mathcal{N}}_{v_j}^l$ sont complémentaire. L'algorithme standard PEG construit à partir du graphe de Tanner est défini [5] :

Pour $j=0$ jusqu'à $n-1$ **faire**

début

Pour $k=0$ jusqu'à $d_{v_j} - 1$ **faire**

début

Si $k=0$

$E_{v_j}^0 \leftarrow \text{edge}(c_i, v_j)$, où $E_{v_j}^0$ est la première branche incidente de v_j et c_i check-node de plus petit degrés courant parmi les check-nodes $E_{v_0} \cup E_{v_1} \cup \dots \cup E_{v_{j-1}}$.

Sinon

parcourir l'arbre du symbole v_j jusqu'à la profondeur l où :

- le cardinal de $\mathcal{N}_{v_j}^l$ n'évolue plus ($|\mathcal{N}_{v_j}^l| = |\mathcal{N}_{v_j}^{l+1}|$), alors $E_{v_j}^k \leftarrow \text{edge}(c_i, v_j)$ où $E_{v_j}^k$ est la k^{eme} branche incidente de v_j et c_i est le check-node de plus petit degrés courant parmi les check-node de $\overline{\mathcal{N}}_{v_j}^l$.
- $\overline{\mathcal{N}}_{v_j}^l \neq \emptyset$ mais $\overline{\mathcal{N}}_{v_j}^{l+1} = \emptyset$, $E_{v_j}^k \leftarrow \text{edge}(c_i, v_j)$ ou c_i est le check-node de plus petit degrés de l'ensemble $\overline{\mathcal{N}}_{v_j}^l$.

fin

fin

Sur la première condition d'arrêt ($|\mathcal{N}_{v_j}^l| = |\mathcal{N}_{v_j}^{l+1}|$) il n'y a pas création de cycle donc pas de possibilité d'avoir des trapping-set aussi mais la deuxième $\overline{\mathcal{N}}_{v_j}^l \neq \emptyset$ et $\overline{\mathcal{N}}_{v_j}^{l+1} = \emptyset$ montre que lors de la connexion nous aurons de nouveau cycle qui seront créés car tous les noeuds de parité sont rencontrés dans l'arbre. Les critères de choix qui minimisent les cycles sont définis dans [5] et [7] et leurs généralisations dans [6]. Pour une variable s_j l'arbre de calcul est modélisé par :

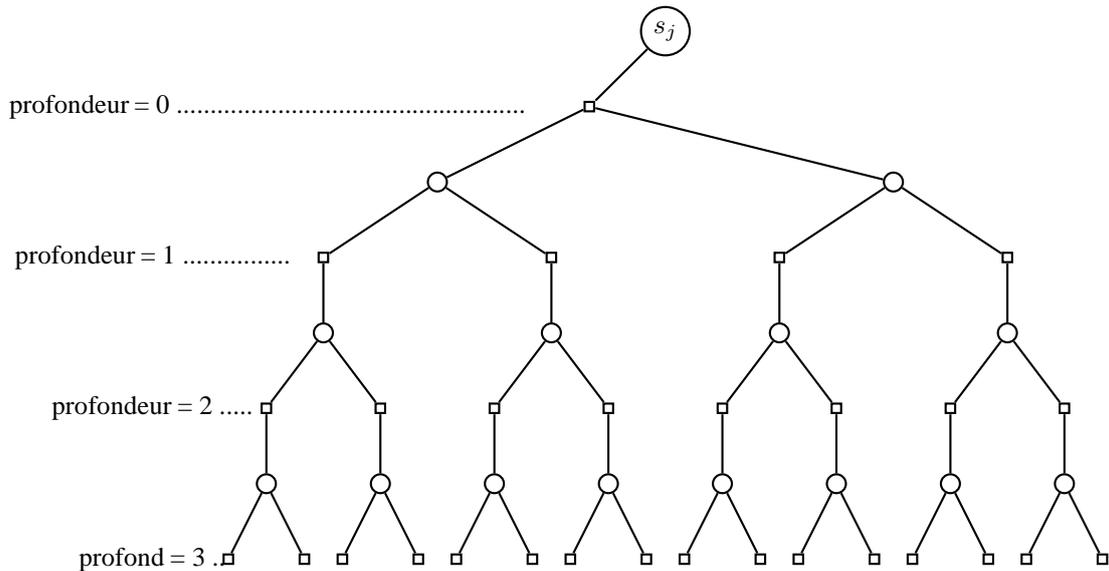


Fig1. Arbre de calcul

3. ELIMINATION TRAPPING-SET TS(5,3)

On appelle trapping-set, noté TS(a,b), un sous-graphe contenant a symbole connectés et b node de parité à connexion impair. D'après Richarson [4], si les *abits* formant le trapping-set sont tous erronés alors le décodeur échoue lors de la tentative de correction. Les critères dans [1],[2] et [3] ne prennent pas en compte les combinaisons qui peuvent exister entre les cycles et ces combinaisons donnent les trapping-set. Dans cette partie

nous décrivons la méthode pour réduire voir éliminer les trapping-set TS(5,3) lors de la conception par la méthode PEG. En se basant sur l'ontologie des trapping-set[10], nous avons 2 modèles possibles pour les trapping TS(5,3), le TS(5,3;8³) et le TS(5,3;6,8,10).

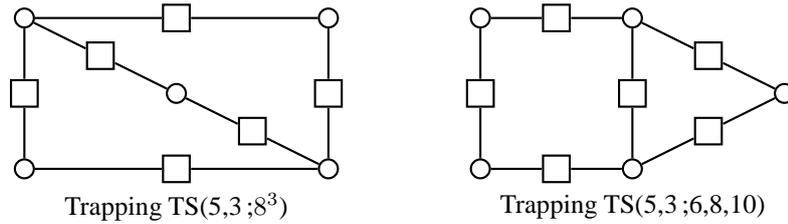


Fig2. Modèle de trapping

– Le TS(5,3;8³) est dû à la création de 2 cycles de taille 8 qui ont 3 symboles en commun.

– Le TS(5,3;6,8,10) est dû à la création d'un cycle d'ordre 6 et d'un cycle d'ordre 8 ayant 2 symboles en commun ou d'un cycle d'ordre 6 et celui d'ordre 10 avec 3 symboles en commun ou d'un cycle d'ordre 8 avec un cycle d'ordre 10 avec 4 symboles en commun.

L'identification des TS(5,3) dans l'arbre pour des variable de degré $d_v = 3$ ce fait par une comparaison de chaque check de l'arbre avec les autres check dont leur apparition est ≥ 2 . Par exemple sur la figure 2 où les noeuds de mêmes couleurs sont identiques décrit quelques scénario de reconnaissance de trapping TS(5,3).

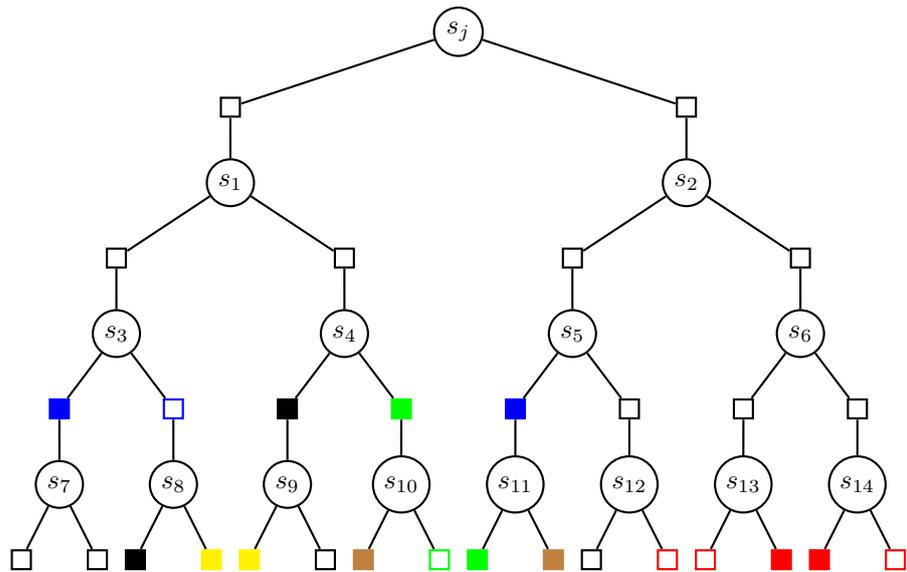


Fig3. Arbre de reconnaissance de trapping TS(5,3)

Les checks donnant des trapping TS(5,3) :

- □ le sous arbre formé par $(s_j, s_1, s_2, s_3$ et $s_5)$ aura 3 neuds de parité impairs qui sont connectés respectivement à $(s_1, s_2,$ et $s_5)$
- □ il suffit de considérer le sous arbre formé par $(s_j, s_2, s_6, s_{13}$ et $s_{14})$ où nous aurons les check relié à $(s_j, s_2, s_{13}$ ou $s_{14})$
- □ en supposant dans ce cas que $(s_1 = s_5,$ et $s_4 = s_{11})$ le sous arbre formé par $(s_j, s_2, s_1, s_{10}$ et $s_4)$

Pour réduire les trapping TS(5,3), il faut lors du parcours de l'arbre repéré les noeuds de parités pouvant donner ces TS(5,3) et les pondérer d'un coefficient. Avec un nouveau critère de selection par ordre de priorité pour les potentiels candidats devient :

- 1) check de profondeur ≥ 2 (pour un girth=6)
- 2) check donnant moins de cycles
- 3) check donnant moins de trapping
- 4) check de plus faible degré

3.1. ALGORITHME PEG-AMELIORE

Nous appelons :

- fircheck d'un noeud c_i , le check de profondeur 0 sur le chemin de v_j à c_i dans l'arbre;
- variable en commun (c_i, c_k) les variables présent sur le chemin de v_j à c_i et le chemin de v_j à c_k
- deph c 'est la profondeur du noeud dans l'arbre
- $2c_i$ signifie deux noeuds identiques présents sur l'arbre à des emplacements différents

L'algorithme PEG-AMELIORE développé, se base sur l'algorithme PEG-Original sauf après parcours si $\overline{\mathcal{N}}_{v_j}^l \neq \emptyset$ et $\overline{\mathcal{N}}_{v_j}^{l+1} = \emptyset$ alors nous évaluons le nombre de Trapping TS(5,3) et nous appliquons le nouveau critère de choix. Par conséquent l'algorithme devient :

Pour $j=0$ jusqu'à $n-1$ **faire**

début(1)

Pour $k=0$ jusqu'à $d_{v_j} - 1$ **faire**

début(2)

Si $k=0$

$E_{v_j}^0 \leftarrow \text{edge}(c_i, v_j)$, où $E_{v_j}^0$ est la première branche incidente de v_j et c_i check-node de plus petit degrés courant parmi les check-nodes $E_{v_0} \cup E_{v_1} \cup \dots \cup E_{v_{j-1}}$.

Sinon

début(3)

tant que $(|\mathcal{N}_{v_j}^l| \neq |\mathcal{N}_{v_j}^{l+1}|)$ et $(\overline{\mathcal{N}}_{v_j}^{l+1} = \emptyset)$ parcourir l'arbre du symbole v_j :

Si $(|\mathcal{N}_{v_j}^l| = |\mathcal{N}_{v_j}^{l+1}|)$, alors $E_{v_j}^k \leftarrow \text{edge}(c_i, v_j)$ où $E_{v_j}^k$ est la k^{eme} branche incidente de v_j et c_i est le check-node de plus petit degrés courant parmi les check-nodes de $\overline{\mathcal{N}}_{v_j}^l$.

sinon $/(N_{v_j}^l \neq \emptyset, N_{v_j}^{l+1} = \emptyset)$

debut(4)

Parcours la liste des noeuds de parité rencontrés sur l'arbre

Si $2c_i$ de deph=3 ont (3 variables en commun) alors incrémente TS(5,3) de c_i

Sinon Si c_i de deph=3, a (3 variables en commun avec $2c_k$ de deph(2,3)) alors incrémente TS(5,3) de c_i

Sinon Si c_i de deph=2, a (3 variable en commun avec c_k de deph=2), (2 variables en commun avec un autre c_k de deph=3) et (que ces 3 noeuds aient même firscheck) alors incrémente TS(5,3) de c_i

Sinon Si c_i de deph=2, a (2 variables en commun et même firscheck avec c_k de deph=1) et aussi (1 variable en commun et firscheck différent avec un autre c_k de deph=2) alors incrémente TS(5,3) de c_i

Sinon Si c_i de deph=2, a (3 variables en commun et même firscheck avec c_k de deph=2) et aussi (1 variable en commun et firscheck différent avec un autre c_k de deph=2) alors incrémente TS(5,3) de c_i

Sinon Si c_i de deph=3, a (4 variables et même firscheck avec c_k de deph=3) et aussi (3 variables en commun et même firscheck avec un autre c_k de deph=2) alors incrémente TS(5,3) de c_i

fin debut(4)

$E_{v_j}^k \leftarrow \text{edge}(c_i, v_j)$ où $E_{v_j}^k$ est la k^{eme} branche incidente de v_j et c_i est le check-node de plus grand profondeur donnant moins de cycle ayant le plus petit nombre de TS(5,3) et plus petit degrés courant parmi les check-node de $N_{v_j}^l$

fin debut(3)

fin debut(2)

fin debut(1)

4. SIMULATIONS ET RESULTATS

Les simulations sont effectuées avec des rendements différents R=0.4 et R=0.5. La réduction des trapping est représentée sur le tableau1.

Rendement (M,N)	R=0.4 (93,155)	R=0.5 (252,504)	R=0.5 (126,252)
TS(5,3) avec PEG ()	21	7	53
TS(5,3) avec PEG Amélioré	17	5	27

La réduction des trapping peut engendrer aussi une réduction des cycles de taille faible.

Rendement (M,N)	R=0.4 (93,155)	R=0.5 (252,504)	R=0.5 (126,252)
cycles(taille 6, taille8) avec PEG	(6,407)	(3,1032)	(3,433)
cycles(taille 6, taille8) avec PEG Amélioré	(4,410)	(5,1005)	(1,417)

Dans chacun des cas nous avons une réduction du nombre de trapping. Cette réduction se traduit par une amélioration des performances du codes

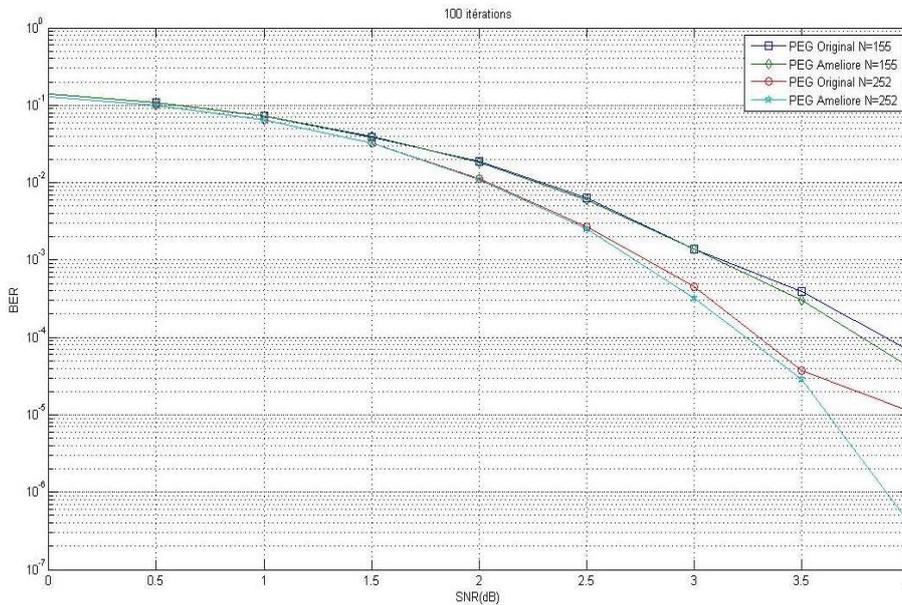


Fig4. Taux d'erreurs binaire en fonction du SNR avec $N=504$, $M=252$

D'après ces résultats nous remarquons que la réduction du nombre de trapping augmente les performances des codes.

5. CONCLUSION

Dans cet article nous avons montré qu'il est possible d'augmenter les performances des codes LDPC basés sur la méthode PEG en évitant non seulement les cycles de petites tailles mais aussi certaines combinaisons interne appelées trapping-set. Nous avons proposé un algorithme permettant de réduire les trapping-set $TS(5,3)$ pour un girth $G \geq 6$. Par des simulations nous constatons que cette réduction de trapping augmente les performances des codes LDPC. Cette réduction de trapping-set pourrait être généralisée par exploitation de l'ontologie des trapping-set et aussi voir leur impact sur les codes quasi-cycliques.

6. Bibliographie

- [1] R. G. GALLAGER, « Low-Density Parity-Check Codes », *M.I.T. Press, ambridge*, 1963.
- [2] D. J. C. MACKAY and, R. M. NEAL « Near Shannon limit performance of low-density parity-check codes », *Electron. Lett*, n° 32, 1645-1646, August 1996.
- [3] T. J. RICHARDSON, M. A. SHOKROLLAHI, and R. L. URBANKE, « Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes », *IEEE Trans on Inf. Theory*, vol. 47, n° 2, pp 619-637, February 2001.
- [4] T. J. RICHARDSON, « Error floors of ldpc codes », in *Proc. 41st annual Allerton Conf. on Communication, Control and Computing*, pp 1426 1435, 2003.

- [5] X.-Y HU, E.ELEFThERIOU, and D.M.ARNOLD, « Regular and Irregular Progressive Edge-Growth Tanner Graphs », *IEEE Trans on Inf. Theory*, vol. 51, n° 1, pp 386 398, January 2005.
- [6] D.DECLERCQ, M.FOSSORIER, B.VASIC, C.POULLIAT, and A.ROUMI, « LDPC Codes : from Basics to advanced topics. What practical applications can learn from theory », Part I :Binary LDPC Codes, June 2010.
- [7] X. HUA, and A.H.BANIHASHEMI, « Improved Progressive-Edge-Growth (PEG) Construction of Irregular LDPC Codes », *IEEE Commun. Letters*, vol. 8, n° 12, pp 715 717, December 2004
- [8] C.T. HEALY and, R. C. DE LAMARE « Decoder-Optimised Progressive Edge Growth Algorithms for the Design of LDPC Codes with Low Error Floors », *IEEE Commun. Lett.*, vol. 16, n° 6, pp. 889-892, Jun. 2012.
- [9] G. RICHTER and, A. HOF « On a construction method of irregular LDPC codes without small stopping sets », *IEEE Int. Conf. Commun.*, n° pp 1119 1124, Istanbul, Turkey, Jun. 2006,
- [10] « [http ://www2.engr.arizona.edu/ vasiclab/Projects/CodingTheory/TrappingSetOntology.html](http://www2.engr.arizona.edu/vasiclab/Projects/CodingTheory/TrappingSetOntology.html) »,

1. Introduction

Software processes are complex and critical entities. There is a constant need for enterprises to represent them and to use them as a basis for processes' stakeholders' communication, training, and know-how capitalization. According to Curtis et al, a software process model (process model, for short), is defined as “*an abstract description of an actual or proposed software process which represents selected process elements that are considered important to the purpose of the model and can be enacted by a human or a machine*” [3]. By process element it is meant, every activity, workproduct, role, agent or tool used during the realization or the maintenance of a piece of software. A workproduct is defined as any artifact consumed, modified, or produced by the activities of the process. For more productivity, a process model can be taken as an input by a process engine for execution. The objectives of such an execution are manifold and include the process workproducts management. However, in the last decade, the software process community has been strongly marked by the advent of the model driven engineering (MDE), which impacted the nature of workproducts in development processes. Indeed, they have noticeably changed in that sense:

- They are essentially models that are defined, refined, transformed, queried and analyzed along the process activities. Hereafter, we use the term model-product to designate any workproduct that comes in the form of a model. For instance, a model-product can be a UML model, a petri net, a state machine, etc.

- They are more and more large-sized. Some models may have up to 160000 model elements [6].

A MDE process involves inevitably model-products which life cycles are strongly related. In fact, the relationships between model-products are numerous and of many kinds. One of such relationships is an inter-model-products relationship that can involve two or more different model-product entities. It consists of a set of model elements which can be shared between several model-products. We call this kind of relationship an *overlap* [11]. Overlap relationships cause consistency management problems occur between model-products, especially in complex projects involving a large number of intertwined model-products. In fact, when working on a model-product during a development process, a developer always deems creating, modifying and deleting model elements of a given model-product without accessing the impacts of his actions on other model-products. However, these model elements can be shared with other model-products of the same process which therefore need to remain synchronized all along the process execution.

As argued earlier, since in MDE many workproducts are model-products, it becomes possible to refine their granularity and therefore to automate an important part of their evolution and consistency management along the process execution. To achieve these objectives, new solutions related to process modeling and execution are necessary. From the *Modeling* point of view, a lack of expressivity is noticed for existing PMLs (Process Modeling Languages). In fact, even if some of them [1, 10] define concepts for modeling some aspects of the overlap relationship, they do not support a full representation of such a relationship as they do not support the expression of the existing variants of overlap relationships which can be total, partial, etc., as defined in [11]. Existing PMLs also do not support the representation of a model-product as a *type* i.e., notion of model typing [12], which is important for overlap relationships definition. This notion is presented further in sections 2 and 4.

From the *Enactment* point of view, there is a need for an optimal model-products rep-

resentation, storage and management solution. Such a solution has to avoid redundant storage of shared model elements and thereby to ensure a dynamic evolution and consistency management for model-products.

This paper therefore proposes to address a problematic that covers both dimensions of software process management, i.e., Modeling and Enactment. It presents our solution for each of such dimensions. The solution comprises a specification of model-products overlap relationships during process modeling; and mechanisms for a dynamic evolution and a systematic consistency check of related model-products during process enactment. The presented solution is not tailored to a given PML but is independent from existing process management approaches. We therefore have concentrated on process elements related to model-products.

The rest of this paper is structured as follows. In section 2 we illustrate the previously presented problem through an example and summarize the needs related to the modeling and enactment dimensions of process management. In section 3 we present the state of the art on the management of process model-products with overlap relationships. Section 4 and 5 present our respective solutions in the both modeling and execution dimensions. Section 6 describes the prototype we developed to illustrate the feasibility of our approach. Section 7 concludes the paper.

2. An Example and Illustrations

The process-example we present here represents a part of a larger process used to develop a video-on-demand (VOD) system presented in [5]. It defines the *Analyse* and *Design* activities involved in iterations. The **Figure 1** shows a schematic view of such a process-example. The *Analyse* activity outputs the *analysis model* model-product which is consumed by the *Design* activity which outputs the *design model* model-product. The two model-products of the VOD system are UML 2 models and are represented in **Figure 2**. The *analysis model* (**Fig. 2.a**) is composed of the use cases and analysis classes of the system. As for the *design model* (**Fig. 2.b**), it contains the design classes and interaction elements of the system.

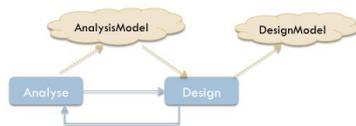


Figure 1. A Schematic View of the Process-Example

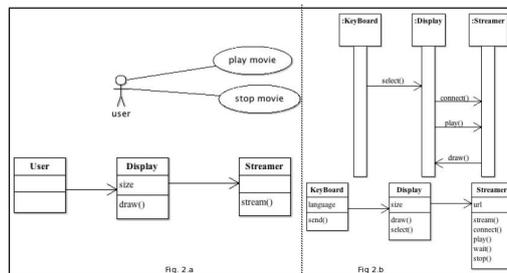


Figure 2. The Model-Products of the VOD System

We consider that two developers called *analyst* and *designer* are (respectively) responsible for these two model-products which share the following model elements: two classes: Display and Streamer, one attribute: size, and two methods: draw() and stream().

During process enactment, from one iteration into another, the *analyst* and the *designer* modify their respective model-products. It is then essential to preserve consistency between those model-products.

Handling the impacts of developer's actions on a given model element depends on the nature of the overlap relationships that involve the considered model-products. An overlap relationship between two or more model-products is characterized by the two following concerns. -**Typing**: it is essential to indicate the type of the shared elements between different model-products in terms of their meta-classes. This defines the type of the overlap. In our example, the type of the overlap between the *analysis model* and *design model* model-products is specified in terms of the Class, Attribute, and Operation UML meta-classes. -**Partial vs. total overlapping**: only some given classes, attributes or operations of the two model-products are concerned but not all of them (e.g. the User and KeyBoard classes are not shared).

In such a situation, given a class, attribute or operation of the *analysis model* model-product, any action of the *analyst* on the corresponding model element has to be presented to the *designer* who gets the option to accept or reject it. The overlap relationship is said to be partial. In the case of a total overlap relationship, such an action would have to automatically be enforced onto the corresponding element of the design model, without any possibility for the *designer* to bypass it. The other model elements of the *analysis model* model-product such as use cases, actors, etc. are not concerned by the previously mentioned constraints.

To achieve this synchronization between model-products, there is a need for an optimal model-products representation, storage and management which fully takes into account their overlap relationships. The model products representation must give access to the constituent model elements and permit the identification and representation of developers' actions on them. As for the storage and management policies, they have to ensure a propagation of developers' actions if they relate to shared model elements.

3. State of the Art

To be effective at enactment time, any information about workproducts interdependency and overlapping needs to be captured at modeling time [3, 7]. Some of the existent process management approaches such as Adele [1], SPEM 2.0 [10] and SPEM4MDE [4] give the possibility to express the overlap relationship during process modeling. Unfortunately, none of them do support a full expression of the above associated characteristics (typing and overlapping nature). The other visited approaches of the literature completely ignore the support of the overlap relationship.

At enactment time, since the two model-products get elements in common, they have to automatically be kept synchronized all along the process execution. Otherwise, an important risk of inconsistency between them is to be faced. Therefore, there is an actual need for the process engine to be provided with mechanisms which are able to enforce a dynamic evolution and consistency management for the two model-products. Such mechanisms have to take into account the formerly mentioned characteristics of the overlap relationship. They have to bring storage solutions for the shared classes, attributes and operations and ensure the management of developers' actions on such model elements during their activities on the two model-products. Unfortunately, in the current process management and enactment solutions, workproducts are considered as a whole and the granularity for ensuring consistency and synchronization between them is the file which

is too coarse-grained.

In the following sections, we present our solution for handling the overlap of model elements between different model-products in both dimensions, i.e. at process modeling time and enactment time. We start by the modeling dimension.

4. Model-Products' Overlap Relationships Modeling

In [3] and [7], authors argue that details on the structure and relationships of any workproduct are related to the process and therefore have to be described during process modeling time. The solution we introduce hereunder is PML-independent and focuses on modeling process model-products. It comes in the form of a meta-model which concepts are related to process model-products and to the overlap relationships between them. The **Figure 3** presents the meta-model. Its main concepts are presented hereafter.

The WorkProduct metaclass is used to represent any artifact produced or consumed by a process activity. ModelProduct is a kind of WorkProduct that represents artifacts that come in form of a model, instance of a metamodel. The ModelType metaclass represents model-product types. Its semantic is inspired from the one defined by Steel et al. in [12]. They define a model type as the set of metaclasses belonging to its metamodel. Therefore, we consider a model-product type as a set of ModelTypeElements. The ModelTypeElement concept specifies any metaclass that belongs to the model-product's metamodel. The motivation of the use of such typing approach is to give the possibility to specify an overlap relationship on the basis of the involved model-products types. In fact, given two or more model-products, they can overlap if and only if they have ModelTypeElements in common. Model-products' types definition during process modeling also give to process engines the possibility to process any model-product with a finer-grained view giving access to its model elements.

The Overlap concept of the metamodel is used to specify overlap relationships between two or more model-products. The existence of such relationship between model-products intrinsically depends on their respective model types. The *specification* of an Overlap corresponds to the intersection of the ModelTypes of the involved WorkProducts. Each common model element is an instance of a metaclass belonging to it. For example, given an instance *ove* of the Overlap metaclass. Its *overlappingWorkProducts* (*oWP*) collection is composed of WorkProducts whose associated ModelTypes own, at least, one ModelTypeElement in common. Thus, at the process enactment time, it is possible for model-products that belong to *oWP* to own common model elements which, in addition, are instances of elements belonging to the *specification* of *ove*. An Overlap is also described by its *nature* attribute. In fact, our proposition refers to Spanoudakis et al. [11] and supports the **total** and **partial** forms of the overlap relationship. As follows is given the semantic associated to each of them.

- **Total Overlap** (OT in the meta-model): all the related model-products share each of their model elements specified by a metaclass belonging to the overlap specification. During a given process modeling, such a kind of relationship gives the possibility to express constraints of this kind: "*Given a set of model-products, each class model element of one of them is a model element of the others too*".

- **Partial Overlap** (OP in the meta-model): the related model-products share some of their model elements specified by meta-classes belonging to the overlap specification. Using such a kind of relationship, one can express a constraint like: "*Given a set of model-*

products, each class model element of one of them may also be a model element of the others too".

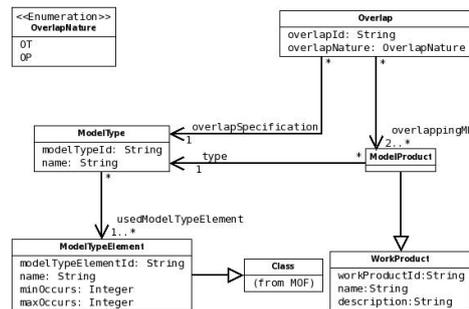


Figure 3. The Meta-Model of Model-Products

5. Overlap Relationship Use During Process Execution

For the running dimension, we have defined and used the notion of software process object (SPO), an entity that represents a process model-product at enactment time. SPOs are used by a process engine to ensure synchronization and consistency between model-products of a running process. In the following, we present their structure, their lifecycle, and show their use during process execution.

When a process engine begins the execution of a process model, it creates the following empty SPOs. (1) A SPO is automatically created and associated to each WorkProduct process model element. (2) For each Overlap relation: (2.1) A SPO (called OverlapSPOR, for Overlap SPO Relation) is created, and (2.2) an observer object is created for each WorkProduct that plays a role in the Overlap relationship. It is associated to the following objects: the SPO associated to the considered ModelProduct, and the current and newly created OverlapSPOR. Observer objects are inspired from the GoF observer design pattern. Their three main roles are resumed as follows: help developers to be informed on the activities of their collaborators on shared model elements, give developers the possibility to approve or reject their collaborators' actions on shared elements, and save actions corresponding to developers' decisions.

A SPO reifies a given model-product. It is used by the process engine to store only model elements which are not shared with other model-products. An OverlapSPOR represents an Overlap relationship. It saves the shared model elements between the model-products involved in the specified overlap relationship. The meta-classes of such model elements belong to the Overlap specification. In fact, many solutions came to us for the storage shared model elements. According to the adopted one, all common model elements are saved in the OverlapSPOR and referenced from the SPOs representing the overlapping model-products. Such a solution imposes the following requirement to be met: *in the case of a partial overlap, when a developer creates, deletes or modifies a model element which meta-class belongs to the overlap specification, there is a need to notify the corresponding actions to all the other developers responsible for model-products involved in the represented overlap relationship.* Observer objects are used in order to satisfy such a constraint.

The initialization of a running process is followed by its main execution stage. During

such a stage, developers execute their tasks in the process and the process engine manages the model-products' evolution using the SPOs. To begin an activity, a developer logs on the process engine which retrieves its needed model-products. Then the developer uses his favorite tools to perform his tasks. At the end of his work, the developer sends his created or modified model-products to the process engine which saves their constituent model elements. Model-products' restoration and storage are realized using SPOs, OverlapSPORs and their associated observers using a protocol we developed in fact.

6. The Approach Prototyping

In order to illustrate the feasibility of our approach, we have developed a prototype that simulates a PSEE (Process-centered Software Engineering Environment) implementing the presented proposition. The prototype has been used to model and execute the process-example presented in section 2 (**Figure 1**). Two main components appear in its architecture: the modeling and the execution ones. The modeling component comes in the form of an Eclipse EMF window. It gives the possibility to a process modeler to create process models in which model-products and their overlap relationships are described using the concepts of the meta-model presented in **Figure 2**. It is based on the meta-model presented in **Figure 2** which is extended by concepts such as Activity, Role, etc. that represent the other constituents of a process. As for the execution component of the prototype, it is responsible for the execution of a process which comprises two main stages as stated in section 5: the SPOs' initial creation and the evolution of their content. Naturally, the content of a SPO evolves when its corresponding model-product evolves. The execution component of the prototype ensures model-products storage and restoration. It is also responsible for the management of the developers' actions on the model elements involved in overlap relationships. In section 1 we identified the need for an optimal model-product representation. Section 5 also highlighted the importance of being able to catch and represent the actions of each developer on model-products. Therefore, adequate representations for model-products and actions on them were necessary for the execution component to be implemented. Recently, operation-based model representation approaches have been developed [8], giving the possibility to represent a model in terms of operations on its model elements. In fact, developers' actions on a given model-product directly come in a form of a set of predefined operations (create, delete, update, etc.). Praxis [2] is one of such model representation approaches. We developed it and used it to represent model-products. Praxis is a framework defining a MOF model representation language. It proposes a model representation inspired from the MOF reflexive API [9] and represents models using six unitary actions.

7. Conclusion and Future Works

This paper has identified the problem of maintaining model-products consistency and synchronization during process execution. Specifically, we have focused on MDE processes where model-products may have overlap relationships between them. Our main objective was to design a solution that gives the possibility to express such kind of relationship and its characteristics during process modeling, but also to use the corresponding information at process enactment time to ensure an optimized model-products consistency and synchronization.

The proposed solution is divided into two parts. In the modeling point of view, we proposed a meta-model which structures the necessary concepts to describe model-products and overlap relationships between them. For the enactment dimension, the solution consisted in defining a set of mechanisms to be used by process engines to ensure synchronization and a dynamic evolution for the process' model-products. The feasibility of the solution is experimented through a case study and a tool we developed. The tool is currently under evaluation by our industrial partner AIRBUS, in the context of the Galaxy research project which this work is part of and that has been co-funded by the French national research agency (ANR).

One perspective of this work is to conduct an empirical study to evaluate the advantages of using our tool in the context of an important software development project. A second one would be to address the *impact* relationship between model-products during process execution. That latter specifies that when a source model-product is modified, a set of operations has to be applied on different target model-products, part of the same impact relationship.

8. References

- [1] BELKHATIR, NOUREDDINE , ESTUBLIER, JACKY andnameMELO, WALCELIO, "THE ADELE-TEMPO experience: an environment to support process modeling and enactment.", *Software Process Modelling and Technology Research Studies Press*, vol. , num. 2007.
- [2] BLANC, XAVIER , MOUNIER, ISABELLE , MOUGENOT, ALIX, MENS, TOM, "Detecting model inconsistency through operation-based model construction", *Proceedings of the 30th international conference on Software engineering, ICSE'08*, vol. num. 2008.
- [3] CURTIS, BILL , KELLNER, MARC I andnameOVER, JIM, "Process modeling", *Communications of the ACM*, vol. 35, num. 9, 1992.
- [4] DIAW, SAMBA , LBATH, REDOUANE , LE THAI, VINH, COULETTE, BERNARD, "SPEM4MDE: a Metamodel for MDE Software Processes Modeling and Enactment.", *3rd Workshop on Model-Driven Tool & Process Integration. Associated to ECMFA, Paris*, vol. num. 2010.
- [5] EGYED, ALEXANDER, "On model typing", *29th International Conference on Software Engineering, ICSE07*, vol. num. 2007.
- [6] EGYED, ALEXANDER, "Automatically Detecting and Tracking Inconsistencies in Software Design Models", *IEEE Trans. Softw. Eng.*, vol. 37, num. 2, March 2011.
- [7] FUGGETTA, ALFONSO, "Software process: a roadmap", *Conference on The Future of Software Engineering*, vol. 97, num. 12, 2000.
- [8] HERRMANNSDOERFER, MARKUS , KOEGEL, MAXIMILIAN, "PEM4MDE: a Metamodel for MDE Software Processes Modeling and Enactment.", *Proceedings of the 1st International Workshop on Model Comparison in Practice, IWMCP 10*, vol. num. 2010.
- [9] OBJECT MANAGEMENT GROUP, "Meta Object Facility (MOF) Core Specification Version 2.0", <http://www.omg.org/spec/MOF/2.0/>, vol. num. 2006.
- [10] OBJECT MANAGEMENT GROUP, "Software & Systems Process Engineering Metamodel (SPEM) Version 2.0", <http://www.omg.org/spec/SPEM/2.0/>, vol. num. 2008.
- [11] SPANOUDAKIS, G. , FINKELSTEIN, A. andnameTILL, D., "Overlaps in Requirements Engineering", *Automated Software Engineering*, vol. 6, num. 2, 1999.
- [12] STEEL, JIM , JÉZÉQUEL, JEAN-MARC, "On model typing", *Software Systems Modeling*, vol. 6, num. 4, 2007.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Evaluation and comparison of RSSI based indoor localization algorithms

Dieng Ndeye Amy*[#] — Chaudet Claude* — Toutain Laurent[#] — Ben Meriem Tayeb*^{**}

* Institut Mines-Télécom / Télécom ParisTech – CNRS LTCI UMR 5141 – Paris, France
{ndieng, chaudet}@telecom-paristech.fr

[#] Institut Mines-Télécom / Télécom Bretagne – Université Européenne de Bretagne – Brest, France
{ndeye.dieng, laurent.toutain}@telecom-bretagne.eu

*^{**} Orange Lab Network Carrier / R&D – Issy-les-Moulineaux, France
tayeb.benmeriem@orange-ftgroup.com

.....

ABSTRACT. A signal transmitted through a wireless channel will typically experience random variations due to the presence of obstruction in the signal path, giving rise to random variations of the received power at a given distance. Thus, a better understanding of techniques used to localize in indoor environment is needed. This paper presents a survey of the most relevant indoor localization techniques, especially those based on RSSI, in addition to strongly experiment RSSI behavior in function of the distance. A comparison results of one-hop localization techniques indicates that in most situations a proposed multipath detection algorithm namely B-MLE globally outperforms the others: the trilateration approach, a random localization, EB-MLE (an extension of B-MLE mechanism) and the maximum likelihood estimation.

RÉSUMÉ. Cet article présente une étude des techniques en matière de positionnement interne, en particulier celles basées sur le RSSI. Une étude expérimentale est également réalisée sur le comportement du RSSI en fonction de la distance. Une comparaison des résultats expérimentaux d’algorithmes de localisation indique que dans la plupart des cas que notre algorithme de détection de trajets multiples à savoir B-MLE donne globalement de meilleurs résultats que les autres: la technique de la trilatération, une localisation aléatoire, EB-MLE (une extension du mécanisme de B-MLE) et l’estimation par maximum de vraisemblance.

KEYWORDS : Range-based, indoor localization, RSSI, log-normal shadow, experimentation

MOTS-CLÉS : Mesure de distance, localisation interne, RSSI, modèle log-normal shadowing, expérience

.....

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

1. Introduction

Geo-localized services are exploding today with the advent of GPS-enabled mobile terminals. Several applications using localization techniques are envisioned in various domains related to the Internet of Things such as personal health monitoring, objects and persons tracking, intelligent homes and cities or environment study and preservation. However, before using localization information in the Internet of Things applications, several issues need to be addressed. First, indoor localization cannot use GPS signals efficiently, due to signal obstruction. Indoor localization may be performed by laser, infra-red or ultrasound-based techniques, which are accurate but require a dedicated and expensive hardware. Radio frequencies-based techniques, on the other hand, do not require specific electronics; as such interfaces are already used for packet communications. In this work, we consider a one-hop Received Signal Strength Indicator (RSSI)-based localization in wireless sensor network with the classic model of the log-normal shadowing. We broadly suppose the presence of sensors nodes with initially unknown location information namely blind target nodes. The location of each blind target node is estimated by using inter-sensor measurements (e.g. distance) and location information of nodes with known location (namely landmarks or references nodes).

We first conduct a thorough analysis of RSSI behavior in indoor environments. We then present an extension of a previously proposed multipath detection algorithm (namely Biased-maximum likelihood algorithm: B-MLE). B-MLE detects only one abnormal landmark in order to improve the accuracy of indoor positioning. Finally, B-MLE is compared to classical trilateration [1], maximum likelihood estimation (MLE) [4], a random localization and other algorithms that detect more than one abnormal landmark.

2. Relationship between measured RSSI and distance

Increasing the received power leads to rising RSSI in non obstructed environment. However in practical indoor scenarios, the ideal distribution of the received power (P_r) is not applicable, because the propagation of the radio signal is interfered with a lot of influencing effects e.g. misplaced wall w.r.t an emitter-receiver communication link, obstructing objects, moving people, reflections on metallic objects. These obstacles can attenuate the signal power between emitter-receiver through scattering, reflection, absorption, which results in shadowing, path loss variation and fading. Thus, the wireless radio channel poses as a severe challenge as a medium for reliable high-speed communication. Statistical models are then used to characterize these variations. The most common model that face these variations is the log-normal shadowing represented by Equation (1) when the emitted power is 0 dBm :

$$PL(d) = PL(d_0) + 10\eta \log_{10}(d/d_0) + \epsilon, \quad \text{and} \quad RSSI = -PL \quad (1)$$

$PL(d_0)$ is a known reference power loss (expressed in dBm) at a reference distance d_0 from the emitter. Usually d_0 is equal to 1 meter in indoor context. $PL(d)$ is the received power loss at the unknown distance d from the emitter to the receiver. η is the path loss exponent and depends on the specific propagation environment. ϵ is a Gaussian zero-mean random variable with standard deviation σ that takes into account the random effects of shadowing and fading. What is required with the use of this model is that, the receiver power at any given distance should be log-normal distributed. The relationship between the measured RSSI values and the path loss is also expressed in Equation (1).

Measurement campaigns are performed to study the behavior of RSSI in indoor environment. We model its variations in different locations in order to define an appropriate indoor localization algorithm. For this study, we use Tmote Sky mote which includes an internal inverted-F and monopole antenna. However, we verified that the radiation pattern of the Tmote Sky antenna is not fully omnidirectional in the plane. In addition, the received signal from a sensor node depends on the receiver antenna orientation. Yang *et al.* [3] confirmed this hypothesis by showing that mounting an antenna on a sensor node deforms its radiation pattern.

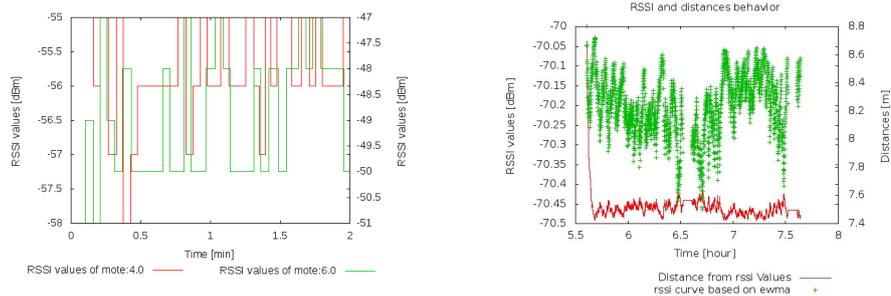
Figure 1 (a) represents RSSI behavior from measurements performed between two receivers that are equidistant from 5.6 m to an emitter. Figure 1 (b) shows for the same room, the RSSI behavior between the same emitter and a third receiver at a distant of 7.3 m and with a constant smoothing factor, $\alpha = 0.05$. We notice from the Figure 1 (b) that the average distance is about 7.6 m, while the actual distance is 7.3 m. This gap distance is due to shadowing effects resulting from obstructing objects in the office, motion from people etc. However, at the same distance between two receivers and the emitter (cf. Figure 1 (a)), we do not report the same RSSI distribution. The use of time series techniques help to better characterize variations observed on measured RSSI values and also to validate the nonconformity of data smoothing methods based on exponential weighted moving average (EWMA). What we also remark is that based on its location, a receiver may measure different behavior on RSSI even if the distance is the same to the emitter, as the propagation condition can differ. This suggest the need to take into account the orientation of the landmarks and receiving target node antennas, interference introduced by the external electronic devices such as the microwave.

3. Algorithms and the System

3.1. Biased-Maximum Likelihood Estimation (B-MLE)

The previous Section 2 highlights the unpredictable behavior of the RSSI in an indoor environment even if the propagation condition remain the same. Hence, the idea to consider the LNSM parameters are based on landmark locations. Furthermore, only 3

Figure 1. RSSI distribution between an emitter and receivers in different configurations



(a) One emitter and two equidistant receivers in same room

(b) One emitter and a receiver in same room; ewma constant smoothing factor $\alpha = 0.05$

landmarks are needed in theory to locate in 2D. In this context, we can afford to rule out the one that has abnormal behavior, and we hope to achieve a coherent set. This dynamic method has been proposed in our previous work [2]. B-MLE approach can be summarized in this way: we identify any abnormal landmark and to exclude it, we replace its measurement by a constant bias. The estimated position and bias term are obtained by examining a global likelihood function, in which we replace the measurements from a landmark \mathcal{O} considered as abnormal by a constant bias term. B-MLE algorithm can be summarized by the equations shown below:

$$L_{\mathcal{O}}(x, y, \beta) = \sum_{\mathcal{L}=1; \mathcal{L} \neq \mathcal{O}}^K L_{\mathcal{L}}(x, y) + L_{\mathcal{O}}(\beta) \quad \text{and} \quad (\hat{x}_{\mathcal{O}}, \hat{y}_{\mathcal{O}}, \hat{\beta}_{\mathcal{O}}) = \arg \max_{x, y, \beta} \{L_{\mathcal{O}}(x, y, \beta)\} \quad (2)$$

Notice that, there is case in which no landmark is considered abnormal which gives $(K + 1)$ likelihood values to compare by selecting the situation that gives the best likelihood value. This identifies the most aberrant landmark, or concludes that the best coherence is achieved when no landmark is ruled out. It is most likely that in some situations, accuracy can be further improved by ruling out more than one landmark.

3.2. Extension of B-MLE (EB-MLE) : Excluding multiple landmarks

As mentioned above, there are cases in which we need to exclude more than one reference node. In this extension of B-MLE, we first study the provided accuracy when excluding two references nodes, then three. Our aim is to identify which method leads to the best positioning accuracy regarding the indoor environment. When more than one reference node is excluded, each set of measurements is replaced by a distinct bias. This allows to keep the same maximization method as above, separating variables for the optimization. For example, in the case we consider two reference nodes, \mathcal{O}_1 and \mathcal{O}_2 , for

exclusion, the global likelihood function needs $1 + K + \binom{n}{2}$ operations and is expressed as:

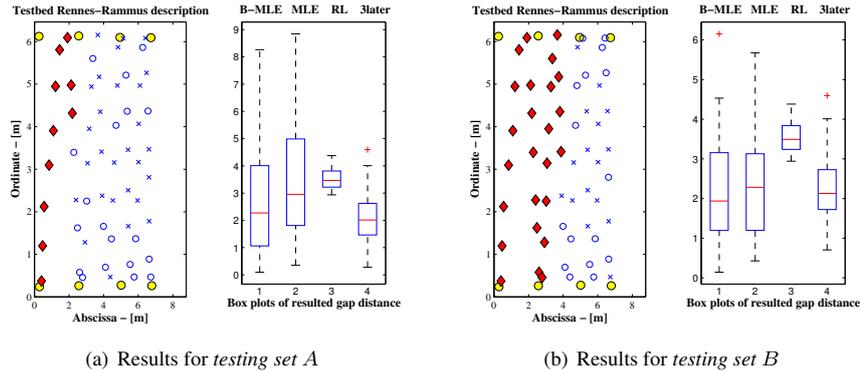
$$L_{\mathcal{O}}(x, y, \beta_1, \beta_2) = \sum_{\mathcal{L}=1; \mathcal{L} \neq \mathcal{O}_1; \mathcal{L} \neq \mathcal{O}_2}^K L_{\mathcal{L}}(x, y) + L_{\mathcal{O}_1}(\beta_1) + L_{\mathcal{O}_2}(\beta_2) \quad (3)$$

4. Results

Algorithms are evaluated in an experimental testbed which is a classroom of size 8.77 m by 6.46 m. This testbed namely Rammus-Rennes is part of the Rennes Rammus platform hosted by the RSM department at Telecom Bretagne. There are 8 landmarks (references nodes) that broadcast 100 beacon packets at a rate of one packet per second and then stop emitting until another request is received from a blind mobile target node. The mobile node measure the RSSI of the beacon frames it effectively receives and calculates its position from this dataset. Each sensor node is a Tmote Sky mote and is placed at a 1.25 m height from the ground which correspond to the hip of a human. However, we did not try to have an isolated room even if the Tmote Sky mote communicates in the opened 2.4 GHz band. A representation of the testbed is reported on the Figure 2. The blue crosses represent the 57 successive positions at which we estimate the mobile node's location. Landmarks are represented by yellow circular dots and the 6th one is mobile. The parameter of the mean/median/standard deviation of the blind mobile target node coordinates are estimated from localization based on MLE [4], trilateration [1], a random localization, the proposed B-MLE algorithm and its extension. We use a random localization performed on the whole testbed to evaluate the contribution of the RSSI. In random localization process, Monte Carlo algorithm is used to draw random coordinates, according to a uniform distribution. Furthermore, these aforementioned algorithms are compared on the sole position estimation by separating each online experiment in two steps. In the first step, we only estimate the values of the propagation model parameters ($PL(d_0)$, η and σ) for each landmark using least square estimation. We use, for this first phase, a set of measurements coming from a subset of positions that we call the *learning set*. Then, on the remaining position (namely the *testing set*), we consider these parameters as constant and compare the three algorithms. We also verified that, the propagation model parameters such as $PL(d_0)$, η and also σ are dependent to the landmarks locations and then presents different values for each landmark.

Two scenarios of *learning set* are considered when comparing these algorithms : scenario *A* first considers the use of the 10 first positions to learn the propagation model parameters and then utilizes the rest of the positions (*testing set*) to test the accuracy of each algorithm. Scenario *B* on the other hand performs the same method but uses the 25 first positions as *learning set*. The experimental results obtained when running the previous mentioned localization algorithms are presented in Figures 2 and 3 for the two

Figure 2. Measurements positions and boxplots of the positioning error in Rennes-Rammus testbed



scenarios *A* and *B*. The blue circles on the left graphs in Figures 2 (a) and (b) are the computed locations at less than 2 m from the true position. Furthermore, the right part in each of both Figures 2 (a) and (b) corresponds to the box-plots of resulted gap distance on the *testing set* and the horizontal line inside each box is the median distance. Observing the results from *learning set A*, trilateration globally performs better than the B-MLE algorithm which outperforms the unbiased MLE and a random localization. To obtain a better sense of the consistency of positioning accuracy, Figures 2 (b) and 3 (b) show the performance of the three algorithms' estimation errors on the same testbed for the scenario *B*. Table 1 and the mentioned figures show that the biased-maximum likelihood algorithm outperforms trilateration, unbiased maximum likelihood and a random localization algorithms. More specially, the average estimation error of the B-MLE algorithm is less than those of trilateration, unbiased MLE and a random localization. However, with the *learning set B*, there are some sole *testing* positions in which B-MLE does not perform better. This is the case for *testing* positions #6 and #23 of the blind mobile target node with *learning set B*. Overall localization results show the performance of B-MLE to increase with the increasing *learning* locations in the *learning set* (no change to other environmental parameters). The performance of trilateration and unbiased MLE on the other hand decreases when the *learning* locations are increasing in the *learning set*.

As we need at least three landmarks to localize and given the number of landmarks in our testbed, we will be able to present the results when ruling out more than one landmark. This comparison is somehow needed to state the performance of B-MLE. The left part of the table 2 presents the corresponding average results when ruling out two landmarks (namely 2B-MLE) in testbed Rammus-Rennes. In this situation, compared to B-MLE, 2B-MLE does not improve the accuracy with both of the *learning sets*. Hence, there is no interest in ruling out more than one landmark. Finally, the right part of the table 2 shows for the same testbed, the obtained results when three landmarks are ruled out (namely 3B-MLE). Ruling out three landmarks (i.e 3B-MLE algorithm) performs better

Figure 3. Results of B-MLE, MLE, trilateration (i.e 3later) and RL in Rennes-Rammus testbed

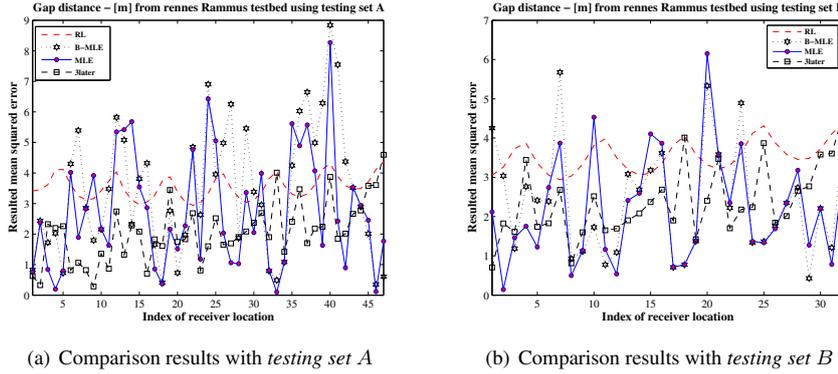


Table 1. Average performance of B-MLE, MLE, 3later and RL from Rammus-Rennes testbed

	Database	Mean distance error (MDE)	Median	STD
Random	A	3.53 m	3.47 m	0.38
MLE	A	3.41 m	2.96 m	2.27
3later	A	2.06 m	2.01 m	0.98
B-MLE	A	2.73 m	2.27 m	1.93
Random	B	3.55 m	3.49 m	0.40
MLE	B	2.38 m	2.28 m	1.94
3later	A	2.36 m	2.13 m	0.92
B-MLE	B	2.20 m	1.94 m	1.40

Table 2. Results from Rennes-Rammus testbed: Average performance of 2B-MLE, MLE, trilateration (i.e 3later) and RL (Left table) and of 3B-MLE, MLE, 3later and RL (right table)

	Dataset	MDE	Median	STD
Random	A	3.53 m	3.47 m	0.38
MLE	A	3.41 m	2.96 m	2.14
3later	A	2.06 m	2.01 m	0.98
2B-MLE	A	2.72 m	2.67 m	1.59
Random	B	3.55 m	3.49 m	0.40
MLE	B	2.38 m	2.28 m	1.40
3later	A	2.36 m	2.13 m	0.92
2B-MLE	B	2.73 m	2.51 m	1.49

	Dataset	MDE	Median	STD
Random	A	3.53 m	3.47 m	0.39
MLE	A	3.41 m	2.96 m	2.14
3later	A	2.06 m	2.01 m	0.98
3B-MLE	A	2.72 m	2.42 m	1.68
Random	B	3.55 m	3.49 m	0.40
MLE	B	2.38 m	2.28 m	1.40
3later	A	2.36 m	2.13 m	0.92
3B-MLE	B	2.72 m	2.42 m	1.43

Dieng *et al.*

than ruling out two landmarks (i.e 2B-MLE algorithm) when the *learning set A* is used. However, with learning set *B*, 2B-MLE performance is comparable to 3B-MLE. In other terms, the comparison of all localization precision indicators (resulted mean distance error (i.e MDE), median and standard deviation) for both scenarios *A* and *B*, emphasizes that excluding more than one landmark is too much. In fact, the performance acquired by B-MLE is better in term of localization precision. In addition, B-MLE requires $K + 1$ maximization operations, while 2B-MLE requires $1 + K + \binom{n}{2}$ such operations.

5. Conclusion

In this research paper, we perform a comparison of a proposed multipath detection algorithm (namely B-MLE) with known localization approaches such as trilateration, maximum likelihood estimation. These algorithms are additionally compared to a random localization process to evaluate the contribution of the RSSI measure. We also conduct a thorough analysis of RSSI behavior in indoor. From our results, it appears that trilateration performs better than MLE. Furthermore, the proposed biased maximum likelihood approach that alleviates multipath effects in indoor environment performs worst than trilateration when the number of the *learning* locations used to compute the propagation model parameters is minimum. However, this situation is not true for all tested positions of the blind mobile target node. Our results also indicate that by increasing the number of samples in the *learning set* B-MLE outperforms the other three algorithms. Besides, we analyze an extended version of B-MLE which rule out more than one landmark. However, this extension of B-MLE is not well adapted to indoor localization.

Future research is needed to explore results of this paper on several different indoor testbeds. As, network self-organization is often realized through "clustering" of the nodes; and localization share several characteristics with "clustering"; this last can be enhanced to provide a faster identification of closed nodes. Global IP prefixes allocation scheme can also be improved as the cluster ID and location together form a stable information, shared by close nodes or nodes moving together.

6. References

- [1] E. Cassano, F. Florio, F. De Rango, and S. Marano. A performance comparison between ROC-RSSI and trilateration localization techniques for WPAN sensor networks in a real outdoor testbed. In *WTS'09*, Prague, Czech Republic, 2009.
- [2] N. A. Dieng, M. Charbit, C. Chaudet, L. Toutain, and T. Ben Meriem. A multi-path data exclusion model for rssi-based indoor localization. In *WPMC 2012*, pages 336–340, Taipei, Taiwan, Sept. 2012. IEEE.
- [3] S. Yang and H. Cha. An empirical study of antenna characteristics toward RF-based localization for IEEE 802.15.4 sensor nodes. In *EWSN'07*, Delft, The Netherlands, 2007.
- [4] G. Zanca, F. Zorzi, A. Zanella, and M. Zorzi. Experimental comparison of rssi-based localization algorithms for indoor wireless sensor networks. In *REALWSN '08*, Glasgow, Scotland, Mar. 2008.

CNRIA'2013

.....

Architecture d'un système multi-agents sémantique : Application au domaine changement climatique et vulnérabilité urbaine.

Ibrahima DIOP^{1,3}, Moussa LO^{1,3}, Jean Marie DEMBELE^{1,3}, Papa Alioune CISSE¹

¹Laboratoire d'Analyse Numérique et d'Informatique (LANI),

UMMISCO UGB – Equipe LIRIMA UGB

³Projet CLUVA (CLimate change and Urban Vulnerability in Africa)

Université Gaston Berger de Saint-Louis - BP 234, Sénégal

E-mail: ibadopfr@yahoo.fr, moussa.lo@ugb.edu.sn, jmdembele@gmail.com,
papaaliounecisse@yahoo.fr

.....

RÉSUMÉ. Dans cet article, nous proposons une architecture de système multi-agents sémantique dans laquelle les agents peuvent accéder aux connaissances de leur domaine et utiliser ces connaissances dans leurs comportements. Nous appliquons ces travaux dans le cadre du projet CLUVA (CLimate change and Urban Vulnerability in Africa). Nous avons précédemment construit une ontologie, une base de règles et une base de faits décrivant les connaissances générales du domaine du changement climatique et vulnérabilité urbaine. Pour l'implémentation, nous utilisons JADE (Java Agent DEvelopment framework) et CORESE (COncceptual REsource Search Engine).

ABSTRACT. In this paper, we propose a multi-agent architecture in which agents can access of its domain knowledge and use this knowledge in their behavior. We apply this work in the project CLUVA (Climate Change and Urban Vulnerability in Africa). We previously constructed an ontology, a rules base and facts base describing the general knowledge of climate change and urban vulnerability. For implementation, we use JADE (Java Agent DEvelopment framework) and CORESE (COncceptual REsource Search Engine).

MOTS-CLÉS : Systèmes multi-agents sémantique, Agents sémantiques, Système complexe, simulation sémantique, Projet CLUVA.

KEYWORDS: Multi-agent system semantic semantics agents, complex system, simulation semantic CLUVA Project.

.....

Actes du 5e Colloque National sur la Recherche en Informatique et ses Applications – Ziguinchor,
Sénégal, Avril 2013

.....

1. Introduction

Dans le domaine des systèmes multi-agents (SMA), les travaux qui s'intéressent à l'utilisation des ontologies sont nombreux. Mais ils vont surtout vers l'identification des agents (agentification) [3], la modélisation et le déploiement [5] et [8]. En effet, dans les SMA, les ontologies sont plutôt utilisées pour la gestion des agents que pour la représentation des connaissances du domaine auxquelles peuvent accéder les agents pour leur choix et décision. Elles ne permettent pas aussi à un utilisateur, pour la compréhension du résultat de sa simulation, de formuler une demande de visualisation par une requête. Certains travaux abordent la notion d'agents cognitifs [1], [2] et [6]. Un agent cognitif est un agent dont la conception se fonde sur des propriétés que l'on attribue habituellement aux êtres humains. La cognition est ici vue comme un mécanisme général de gestion intelligente de l'information : acquisition de nouvelles informations en provenance de l'environnement, raisonnement, prise de décision, etc. Bien que nommés agents cognitifs, ces agents n'ont pas accès aux connaissances de leur domaine mais plutôt à des informations sur leur environnement.

Dans la réalité, les agents d'un système se reconnaissent et ont besoin d'accéder aux connaissances de leur domaine d'évolution pour prendre des décisions (pour la plupart des cas). Mais aussi la visualisation du résultat de la simulation inclue une intension (une demande ou une requête à formuler), par exemple : voir le comportement des sociétés qui émettent le gaz à effet de serre CH₄ avec un débit d'émission supérieur à 500 tonnes par mois.

Nous tentons de nous rapprocher encore plus de la réalité en permettant aux agents d'accéder aux connaissances de leur domaine pour les doter de comportements sémantiques leur permettant une meilleure communication. Nos travaux permettent aussi d'utiliser les connaissances du domaine pour une visualisation sémantique du résultat de la simulation ; ceci, par une requête sur le résultat de la simulation et les connaissances du domaine. Nous appelons un tel SMA, un SMA sémantique.

Cet article présente d'abord le contexte de nos travaux. Dans la section 2, nous proposons une architecture de SMA sémantique pour répondre au besoin de fournir des connaissances du domaine aux agents qui y évoluent. La section 3 présente une application de cette architecture dans le cadre spécifique du phénomène pollueur payeur du domaine du changement climatique et de la vulnérabilité urbaine.

1. Contexte

Le projet CLUVA (www.cluva.eu), soutenu par l'Union Européenne dans le cadre du FP7-ENV-2010, a pour objectif de développer des méthodes et des connaissances pour la gestion des risques climatiques, la réduction de la vulnérabilité, l'amélioration

de la capacité d'adaptation et de la résilience des villes africaines face aux changements climatiques.

Pour cela, l'équipe CLUVA de l'Université Gaston Berger de Saint-Louis développe une plateforme, nommée *Intelligences Saint-Louis*, de gestion des connaissances du domaine de la vulnérabilité urbaine face aux changements climatiques pour la ville de Saint-Louis. Cette plateforme admet un module de simulation sémantique basé sur l'architecture du système multi-agents sémantique proposé ci-dessous.

Pour illustrer notre approche dans cet article, nous allons prendre la théorie du pollueur payeur, qui plutôt que de sanctionner les gros pollueurs signataires du protocole de Kyoto, les fait tabler sur des principes régulant l'économie de marché pour encourager la réduction des émissions des gaz à effet de serre. Dans les faits, cela consiste à mettre fin à la gratuité de l'émission de ces gaz, qui est désormais réglée par un système de quotas par tonne, ou "permis de polluer", que l'on achète, selon les lois de l'offre et de la demande.

Dans la section qui suit, nous présentons une architecture d'un SMA sémantique.

2. Une architecture de système multi-agents sémantique

Dans cette architecture de système multi-agents, nous avons un agent particulier appelé ciel sémantique et un ensemble d'agents sémantiques exploitant les connaissances du domaine dans leurs comportements (voir la figure 1).

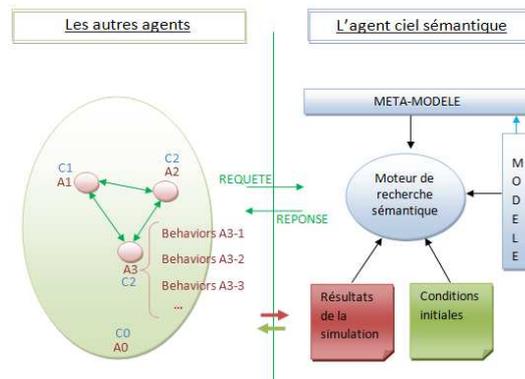


Figure 1: Architecture d'un système multi-agents sémantique

a) Agent ciel sémantique

L'agent ciel sémantique est fondamental dans l'architecture. C'est le premier agent créé et qui se charge d'initialiser les autres agents, de donner les connaissances aux agents qui les demandent et de récupérer et de stocker les résultats de la simulation.

Diop et al.

Dans cette partie, nous faisons un arrêt sur les cinq éléments qui composent l'agent ciel sémantique à savoir :

- **Méta-modèle** : Il contient les connaissances générales du domaine représentées par des ontologies (concepts, attributs, relations), des règles et des faits. Dans le cadre du projet CLUVA, les connaissances générales sont représentées par l'ontologie OntoCLUVA [7] et par les bases de règles et de faits que nous avons établis dans des travaux antérieurs.

- **Modèle** : Il vient compléter le méta-modèle par des connaissances spécifiques à un phénomène à étudier (ex. théorie du pollueur payeur). Pour cela, il contient des concepts et des relations spécialisant ceux du méta-modèle et ajoute des attributs aux concepts du méta-modèle. Il contient aussi les règles et les faits spécifiques au phénomène. De ce fait, le modèle est lié au méta-modèle.

- **Conditions initiales** : Elles sont représentées par des faits en RDF contenant les informations initiales sur les agents sémantiques à créer.

- **Résultats de la simulation** : Ils sont représentés comme les conditions initiales par des faits en RDF, mais contenant les informations sur l'état et les transactions de chaque agent pour chaque pas de temps. Ces informations sont obtenues de la part des agents sémantiques initialement créés qui envoient pour chaque pas de temps leur situation à l'agent ciel sémantique qui les écrit dans le fichier RDF résultats de la simulation.

- **Moteur d'inférence et de recherche sémantique**: permet de conduire des raisonnements logiques et de dériver des conclusions à partir des connaissances. Ce moteur permet aussi la recherche sémantique. Dans notre cas, nous utilisons le moteur d'inférence et de recherche CORESE [9] où nous chargeons le Méta-modèle, le modèle, les conditions initiales de la simulation et les résultats après simulation.

b) Les agents sémantiques

Un agent sémantique est un agent associé à un concept de l'ontologie du domaine et qui a des comportements sémantiques. Un comportement sémantique est un comportement qui peut accéder aux connaissances du domaine et les utiliser. Cela rend la communication entre agents sémantiques plus riche.

c) Communications entre l'agent ciel sémantique et les agents sémantiques

Il y a deux types de communications entre l'agent ciel sémantique et les agents sémantiques.

Les agents sémantiques, après leur création et pour chaque pas de temps ou chaque événement, envoient des messages à l'agent ciel sémantique pour lui rendre compte de leur état. Ainsi l'agent ciel sémantique écrit le contenu de ces messages dans une base de faits qui constitue le résultat de la simulation.

Ils peuvent aussi lui envoyer des messages dont les contenus sont des requêtes en SPARQL pour demander des connaissances et l'agent ciel sémantique répond par des messages dont les contenus sont des faits en RDF correspondant aux connaissances demandées.

Pour visualiser les résultats de simulations, l'agent ciel sémantique charge dans son moteur d'inférence la base de faits résultats de la simulation générée. Ainsi il peut recevoir et traiter des requêtes de visualisation.

La section suivante applique cette architecture pour un système multi-agents représentant la théorie du pollueur payeur dans le cadre du projet CLUVA.

3. Application dans le cadre du projet CLUVA pour la mise en place du système multi-agents sémantique pollueur payeur

L'approche de SMA sémantique que nous avons proposé dans la partie précédente apporte des solutions aux problèmes posés par les SMA classiques, en permettant : une plus riche communication entre les agents ; un accès des agents aux connaissances du domaine dans leur processus de décision ; une visualisation sémantique du résultat de la simulation. Nous appliquons cette approche pour le système pollueur payeur.

3.1 Agent ciel sémantique pour le phénomène pollueur payeur

Après sa création, l'agent ciel sémantique du pollueur payeur charge d'abord le méta-modèle contenant les connaissances générales du domaine du changement climatique et de la vulnérabilité urbaine. Il charge ensuite le modèle contenant les connaissances spécifiques liées au cadre d'application Pollueur-Payeur. Ces deux modèles de connaissances (méta-modèle et modèle) sont décrites par des ontologies, des règles et des faits (comme illustré dans la figure 2). Aussi, faut-il noter que ces deux modèles entretiennent des relations car le modèle spécifique ajoute des concepts et des relations spécialisant ceux du méta-modèle. Il ajoute aussi des attributs aux concepts du méta-modèle.

Pour initialiser les autres agents appelés agents sémantiques, l'agent ciel sémantique charge les conditions initiales qui sont des faits en RDF.

Après leur initialisation, les agents sémantiques évoluent dans le temps en utilisant les connaissances du domaine dans leurs comportements ce qui rend plus riche leurs communications. Dans leur évolution, les agents sémantiques rendent compte de leurs activités pour chaque pas de temps à l'agent ciel sémantique qui traduit les contenus de ces messages de compte rendu en annotations RDF qu'il écrit dans une base d'annotations nommée résultat de la simulation.

La figure 2 suivante présente l'agent ciel sémantique proposée pour étudier la théorie du pollueur payeur. Dans cet agent, nous montrons plus en détail la composition

Diop et al.

du méta-modèle, du modèle, de la relation entre ces deux modèles. Mais aussi dans cette architecture nous donnons l'ordre du chargement des connaissances (méta-modèle et modèle), des conditions initiales et éventuellement du résultat de la simulation pour la visualisation.

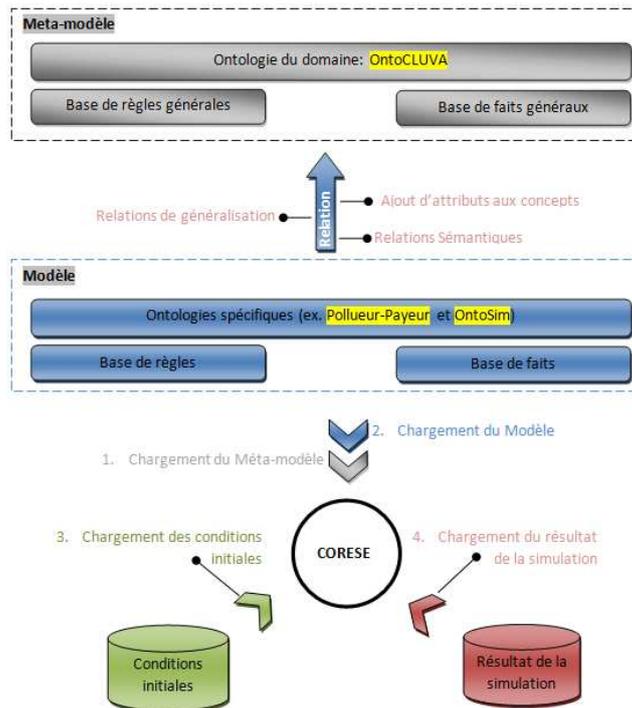


Figure 2: Agent ciel sémantique du phénomène pollueur payeur

3.2. Les agents sémantiques et leurs communications

Dans la définition des agents sémantiques, on déclare un attribut constant contenant le concept de l'ontologie associé à ces agents. Cet attribut permet de savoir qu'un tel agent est associé à un tel concept du domaine.

Cela aide dans la communication entre les agents sémantiques. Ainsi un agent sémantique peut envoyer un message groupé à une catégorie d'agents sémantique (par exemple, envoyer un message à tous les agents associés au concept <http://www.cluva.eu/pollueur-payeur#Societe>).

Le fait que le résultat de la simulation soit au format RDF et que dans la définition d'un agent sémantique on ait un attribut constant contenant le concept de l'ontologie

associé à cet agent aide à la compréhension du résultat de la simulation. Ainsi, ils permettent de faire des filtres dans la visualisation du résultat, par exemple :

- Visualiser les agents pollueurs (c'est-à-dire les agents associés au concept <http://www.cluva.eu/ontocluva#Pollueur>) ;
- Visualiser les agents pollueurs qui émettent du CO₂ ;
- Visualiser les laitiers qui émettent un Gaz à effet de serre ;
- Visualiser les sociétés qui achètent du crédit d'émission ;
- Visualiser les sociétés qui ont une amende.

Ainsi, l'architecture est très ouverte. Elle demande ce qu'on veut visualiser (l'intension de visualisation) dans le résultat. Aussi, dans la visualisation, le moteur sera en mesure de faire des inférences.

4. Conclusion

Dans cet article, nous avons proposés une architecture multi-agents dans laquelle les agents peuvent accéder aux connaissances du domaine et utiliser ces connaissances dans leurs comportements.

Pour cela, dans notre architecture, nous avons créé un agent particulier, nommé ciel sémantique. C'est dans cet agent que nous avons chargé les connaissances du domaine décrites par des ontologies, des règles et des faits. Cet agent communique avec les autres agents qui veulent accéder aux connaissances.

Cette communication se fait par envoi de messages. Les agents désireux de connaissances envoient des messages dont les contenus sont des textes correspondant à des requêtes SPARQL, à l'agent ciel sémantique. Ce dernier, accédant aux connaissances, traite les demandes et donne des retours aux envoyeurs.

Nous appliquons ces travaux dans le cadre du projet CLUVA. Nous avons précédemment représenté, dans le cadre de ce projet CLUVA, les connaissances du domaine du changement climatique et vulnérabilité urbaine, par une ontologie (nommée OntoCLUVA), des règles et des faits.

Pour la programmation multi-agents, nous utilisons JADE [4] (Java Agent DEvelopment Framework) qui est un intergiciel pour le développement d'applications pair à pair d'agents et pour la programmation sémantique, nous avons utilisé l'API CORESE (COncceptual REsource Search Engine) qui est un moteur d'inférence et de recherche sémantique.

Nous avons fait une première implémentation de cette architecture avec un simple système multi-agents sémantique composé d'un agent sémantique, qui envoie une demande de connaissance à un agent ciel sémantique qui après chargement des

connaissances répond à cette demande. Ce qui permet à l'agent sémantique d'envoyer un message à d'autres agents sémantiques en se basant sur les connaissances reçues.

La suite du travail d'implémentation de cette architecture permettra :

- la création d'un SMA sémantique à partir de conditions initiales données à un agent ciel sémantique,
- l'envoi des messages d'états d'agents sémantiques et leur stockage par l'agent ciel sémantique dans le fichier résultats de la simulation,
- et la visualisation d'un SMA sémantique en se base sur les connaissances du domaine, les résultats de la simulation et une requête de visualisation.

5. Références

- [1] A. El Fallah Seghrouchni. Principes et architecture des systèmes multi-agents», Chapitre Les modèles de coordination d'agents cognitifs. Hermès, 2001. 19
- [2] A. Rao et M. Georgeff. «An Abstract Architecture for Rational Agents». Proceedings of C. Rich, W. Swartout, et B. Nebel, the International Workshop on Knowledge Representation, KR'92, pp 439–449, 1992. 11, 20
- [3] Armel Ayimdji, Souleymane Koussoube, Laure Pauline Fotso, Développement des systèmes multi-agents à partir d'ontologies, CARI 2010.
- [4] Bellifemine F., Caire G., Trucco T., Rimassa G., "JADE : Programmer's Guide", Février 2002.
- [5] Csongor I. Nyulas, Martin J. O'Connor, Samson W. Tu, David L. Buckeridge, Anna Okhmatovskaia, Mark A. Musen, "An Ontology-Driven Framework for Deploying JADE Agent Systems," wi-iat, vol. 2, pp.573-577, 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008
- [6] F. Balbo, A. El Fallah Seghrouchni, et S. Pinson. «Organisation et applications des SMA», Chapitre La coordination d'actions par planification Multi-Agents. Hermès, 2002. 19
- [7] Ibrahima Diop, Moussa Lo, Adrien Coly, Choix méthodologiques pour la construction d'une ontologie du domaine de la vulnérabilité urbaine face aux aléas liés aux changements climatiques, Actes du 4e Colloque National sur la Recherche en Informatique et ses Applications – Thiès, Sénégal, Avril 2012.
- [8] M. Laclavik, Z. Balogh, M. Babik, L. Hluch, AGENTOWL: Semantic knowledge model and agent architecture, Computing and Informatics, Vol. 25, 2006, 419–437.
- [9] Olivier Corby, Rose Dieng-Kuntz, Catherine Faron-Zucker, Fabien Gandon, and Alain Giboin. Le moteur de recherche sémantique Corese. In Proc. of the Workshop Reasonner le web sémantique avec des graphes, AFIA platform, Nice, May 2005.

Traitement décentralisé de requêtes de biodiversité

Ndiouma Bame¹, Hubert Naacke², Idrissa Sarr¹ et Samba Ndiaye¹

¹Département de mathématique-informatique

UCAD, Dakar SENEGAL

ndioubame@yahoo.fr

samba.ndiaye@ucad.edu.sn, idrissa.sarr@ucad.edu.sn

²LIP6

UPMC Sorbonne Universités

France

Hubert.Naacke@lip6.fr

RÉSUMÉ.

Le GBIF est un portail mondial pour le partage de bases de données dans le domaine de la biodiversité. Il a pour objectif de fédérer et partager les données de biodiversité existant dans de nombreux laboratoires à travers le monde. Avec un nombre croissant de fournisseurs qui ajoutent de nouvelles données et d'utilisateurs manifestant de nouveaux besoins qui interrogent la base, l'infrastructure est confrontée à des problèmes de disponibilité et d'expressivité limitée des requêtes. Pour faire face à ces problèmes, nous envisageons une solution qui passe à l'échelle avec un coût relativement faible. Dans cette perspective, nous proposons une architecture décentralisée et non intrusive pour interroger les données du GBIF en s'appuyant sur une infrastructure distribuée. Nous définissons une stratégie de répartition dynamique des données, adaptée au contexte du GBIF. Nous démontrons la faisabilité de notre approche par l'implémentation d'un prototype exécutant des requêtes jusqu'ici non supportées par le GBIF.

ABSTRACT.

The GBIF is a global portal for biodiversity databases sharing. Its goal is to federate and share the biodiversity data existing in many laboratories across the world. The global biodiversity data faces two problems, namely the data availability and a poor expressiveness of queries mainly due to a growing number of users, which express more new needs. To deal with these problems, we envision a scalable and relatively low cost solution. With this in mind, we propose a non-invasive and decentralized architecture for processing GBIF queries over a distributed system. We define a dynamic strategy for data distribution that fits the GBIF requirements. We demonstrate the feasibility of our solution by a prototype implementation, which allows for processing extra query types, up to now unsupported by the GBIF portal.

MOTS-CLÉS : masses de données, répllication et distribution de données, cloud computing, traitement des requêtes, GBIF.

KEYWORDS: data replication and distribution, cloud computing, query processing, GBIF.

1. Introduction

Le GBIF est un système d'information qui a pour objectif de fédérer et de partager les données de la biodiversité à l'échelle mondiale [1][3][5]. Sa base de données est complétée continuellement par des correspondants nationaux. Elle contient aujourd'hui près de 400 millions d'enregistrements. Avec un nombre croissant à la fois de fournisseurs qui ajoutent de nouvelles données, et d'utilisateurs qui interrogent la base [2][5], l'infrastructure actuelle du GBIF, qui est centralisée via un portail, peine à servir toutes ces demandes en un temps raisonnable. Ces limites que nous avons recensées dans [17] posent, d'une part, un réel problème de disponibilité des données tout en empêchant un usage réellement interactif de ces données, d'autre part, les infrastructures informatiques sont en pleine évolution : les nuages informatiques (*cloud*) sont omniprésentes et permettent d'accéder à des ressources quasi-illimitées de stockage et de calcul [10]. Ceci incite à concevoir de nouvelles solutions pour la gestion de gros volumes de données, garantissant des accès rapides et un coût relativement abordable en fonction de la charge applicative [16] [8] [9]. Les objectifs de ce travail sont :

1) de décentraliser l'accès au portail à travers plusieurs participants afin de paralléliser les accès et les traitements; 2) de réduire les congestions et d'augmenter la réactivité du système vis-à-vis de ses usagers ; 3) de rapprocher les données des utilisateurs afin de réduire, le plus possible, les accès distants ; 4) de réduire le coût et/ou le temps de réponse des requêtes de l'utilisateur.

Pour ce faire, nous répliquons dynamiquement et partiellement les données du GBIF en fonction des classes d'accès des requêtes et de la localisation des utilisateurs. De fait, les données du GBIF sont répliquées partiellement sur des nœuds locaux (associant des utilisateurs à une base de données locale) en fonction des types de requêtes émises par les utilisateurs se trouvant sur un nœud local. Un fragment du GBIF est créé et répliqué sur un nœud local à chaque fois qu'un utilisateur accède à ce fragment et que ce dernier ne se trouve pas déjà sur la base de données locale associée au nœud en question. Cette stratégie, qui est presque similaire au principe de cache des données en mémoire centrale, a pour avantage de décentraliser les accès aux données surtout pour les requêtes fréquentes et elle permet d'exploiter au mieux les ressources disponibles (GBIF et nœuds locaux). De plus, dès qu'un fragment est présent sur un nœud local, ce dernier est sollicité en remplacement du GBIF à chaque fois qu'une requête veut accéder au fragment. Cette solution est adaptée dans le contexte du GBIF où les données sont rarement modifiées et donc les copies sur les nœuds nationaux correspondent le plus souvent à celles du GBIF. Par ailleurs, les nœuds nationaux sur lesquels nous répliquons partiellement les données du GBIF et que nous appelons par la suite des participants peuvent collaborer pour exécuter des requêtes manipulant des données se trouvant sur un ensemble de participants. Ce mécanisme de collaboration pour traiter des requêtes réparties écarte notre solution des techniques de cache de données classiques.

3 Architecture répartie pour le traitement parallèle de requêtes de biodiversité

La suite de ce papier est organisée comme suit : la section 2 décrit l'architecture décentralisée que nous proposons pour manipuler les données du GBIF ainsi que notre stratégie de répartition des données. La section 3 présente les mécanismes de traitement des requêtes. La section 4 détaille la validation expérimentale de notre proposition et présente les résultats obtenus et les contributions de notre travail. Les sections 5 et 6 contiennent respectivement quelques travaux relatifs au traitement de requêtes dans les environnements distribués et une conclusion et les perspectives de ce papier.

2. Architecture et modèle de données

2.1. Architecture décentralisée pour l'accès aux données du GBIF

Comme le montre la figure 1, notre architecture est composée du portail GBIF, des nœuds participants et des clients.

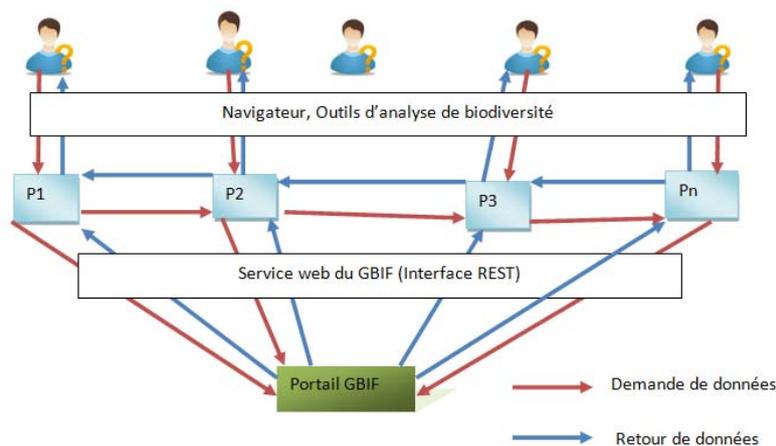


Figure 1: Architecture distribuée pour traiter les requêtes du GBIF

- **Le portail** : c'est un entrepôt récapitulant l'ensemble des données primaires de biodiversité accessibles à travers le réseau GBIF [1]. Cette base est utilisée pour traiter directement les requêtes ou utilisée comme passerelle pour accéder aux bases des fournisseurs. On accède aux données de la base du portail par navigateur web et/ou par des appels web service.
- **Participant (P_i)**: Les différents participants sont reliés par un réseau et collaborent pour échanger des données et traiter des requêtes, grâce au service de requêtes. Un

CNRIA '13

utilisateur accède aux données du GBIF à travers un participant P_i . Un participant, P_i , héberge une base de données locale BD_i et un catalogue. Chaque catalogue local qui contient toutes les informations relatives à la localisation de chaque fragment et de ses répliques à travers le système.

- **Service de requêtes SR** : c'est une couche au dessus des BD locales (BD_i) des participants P_i , qui gère les communications et collaborations entre les différents participants. Il sert d'interface entre le système, les utilisateurs et le portail GBIF. Le service de requête est chargé du choix du site optimal pour l'évaluation des requêtes. Il gère également les transferts de données entre les Participants et les transferts entre le GBIF et les participants.

2.2. Modèle de répartition des données : Fragmentation et réplification dynamiques

Dans notre architecture, nous avons une base de données BD_i locale dans chaque participant P_i . Chaque BD_i dispose du même schéma de données que celui de la base de données complète du GBIF. Ceci permet à tout utilisateur connaissant le schéma de la base du GBIF d'interroger sans limitations les données du GBIF. Par contre, pour les données, seules des portions de la base du GBIF sont répliquées sur les BD_i . Les portions répliquées contiennent les données fréquemment manipulées en lecture, ce qui permet de sauvegarder la bande passante tout en réduisant le temps de réponse [19]. Pour déterminer les données à répliquer, nous nous basons sur les prédicats des requêtes qui sont soumises à partir du participant concerné. Une telle réplification, faite à la demande, permet de minimiser les coûts de stockage des BD_i s et de communication avec le GBIF tout en favorisant le parallélisme des accès et des traitements. De plus cette stratégie de réplification permet de placer les données aux endroits où elles sont fréquemment utilisées. Une réplification totale qui consiste à copier complètement la base de données centrale du GBIF dans chaque participant, affecterait négativement le coût de traitement des requêtes vu le volume de la base. A cela, s'ajoute le fait qu'une réplification totale ferait que certaines données dans les participants n'y seraient jamais exploitées. C'est pourquoi nous adoptons une réplification partielle dynamique en se basant sur les hypothèses selon lesquelles, les tailles des bases de données locales des participants sont limitées et inférieures à celle de la base du GBIF et que les coûts de transfert d'une même donnée entre participants (se situant dans un cloud ou un même réseau local) est négligeable devant le coût de transfert de la même donnée entre le GBIF et un participant. Nous avons choisi une fragmentation horizontale dérivée en utilisant le nom de l'espèce adressée par la requête. La fragmentation est faite à la demande en utilisant les prédicats d'une ou d'un ensemble de requêtes. Lorsqu'un fragment f qui existe dans un participant P_i , est très sollicitée par les utilisateurs U_j d'un participant P_j , alors f est répliquée une nouvelle fois depuis P_i vers le participant P_j . Cela permettra aux utilisateurs U_j d'accéder localement aux données qu'ils demandent. Cette re-réplification résulte d'une collaboration entre les participants.

3. Traitement des requêtes sur le GBIF

Dans cette section, nous décrivons le mécanisme d'exécution des requêtes dans le système. Selon la localisation des fragments nécessaires au traitement d'une requête, le mécanisme d'exécution utilisé diffère. Pour aborder ces mécanismes, nous avons posé les hypothèses suivantes : (i) ; le nombre de fragments impliqués par les différentes requêtes du même utilisateur est limité. En d'autres termes une requête concerne seulement une petite portion de la base de données du GBIF, (ii); toutes les données nécessaires à l'exécution d'une requête peuvent tenir dans un participant tant pour le coût de stockage que pour le coût de traitement, (iii) ; pour prendre en compte la possibilité de traiter une requête en parallèle, sans toutefois disperser le traitement sur un trop grand nombre de participants, nous nous efforçons de limiter le degré de parallélisme au minimum nécessaire. En effet l'accès à un participant induit un surcoût (e.g., ouverture de la connexion et gestion des pannes) proportionnel au nombre de participants. Pour ces raisons, nous considérons que le coût d'accès à deux fragments A et B provenant de deux participants distincts, est supérieur au coût d'accès à A et B provenant d'un seul et même participant. (iv) Les traitements locaux sans nécessité de transfert de données, sont moins coûteux que les traitements répartis pour les mêmes requêtes. Nous favorisons les traitements locaux tout en exploitant les possibilités de parallélisme si nécessaire en fonction de la localisation des données. En effet, la réduction du nombre de sites qui interviennent au traitement d'une requête permet de diminuer les risques de blocage de l'exécution et permet de réduire le temps de réponse avec la diminution des temps de consolidation de résultats. Nous faisons la distinction entre deux types de requêtes ; les requêtes décomposables que l'on peut découper en un ensemble de sous-requêtes qui peuvent être traitées simultanément sur des données réparties à travers différents participants. En opposition, nous avons les requêtes non-décomposables dont l'exécution nécessite que toutes les données impliquées soient dans un participant.

3.1. Traitement de requêtes décomposables :

Une requête décomposable peut être reformulée comme l'union ou l'intersection de plusieurs sous-requêtes accédant chacune à des données disjointes. Lorsque les données sont réparties sur plusieurs participants, la requête décomposable est découpée en un ensemble de sous-requêtes, et chaque sous-requête est envoyée au participant contenant les données concernées. Ce mécanisme qui est similaire aux principes de MapReduce [8][12], permet de paralléliser les traitements en les découpant en des sous-traitements, et minimise ainsi le temps de réponse. Un fragment de donnée peut être répliqué à travers plusieurs participants. Le choix de la réplique à lire pour le traitement d'une requête est guidé par la minimisation des coûts de communication en favorisant les traitements locaux. Pour cela, lorsqu'un fragment F_i est répliqué sur deux participants P_i et P_j , on choisit de solliciter le participant disposant du plus grand nombre de fragments

nécessaires au traitement de la requête. Dans le cas particulier où P_i et P_j ont le même nombre de fragments impliqués dans le traitement d'une requête, alors le participant le moins chargé est choisi.

3.2. Traitement de requêtes non-décomposables

Une requête non décomposable ne peut pas être reformulée comme l'union ou l'intersection de plusieurs sous requêtes, car elle nécessite un traitement final plus complexe qu'une simple opération d'union ou d'intersection, par exemple un regroupement (group by) suivi d'agrégations (min, max). L'objectif visé ici est de traiter une requête quelle que soit sa complexité tout en minimisant le nombre de participants intervenant dans le traitement de la requête pour les mêmes raisons que dans le cas du traitement de requêtes décomposables (cf. section 3.1). Pour cela, on choisit les sites qui ont le plus de données nécessaires à l'exécution de la requête. Parmi ces derniers, on désigne le site disposant de la plus grande quantité de données, comme étant le site d'exécution de la requête. Les autres participants évaluent localement les éventuels prédicats de sélection de la requête, puis transfèrent ensuite leurs réponses vers le participant désigné. La requête est finalement évaluée sur le participant désigné.

4. Validation expérimentale

Dans cette section, nous évaluons la faisabilité et les performances de notre solution que nous allons ensuite comparer avec une solution qui préconiserait une réplification totale des données du GBIF sur chaque participant sans que ces derniers ne collaborent. L'objectif de notre expérimentation est de montrer que notre solution est faisable avec des performances raisonnables et est meilleure qu'une solution de réplification complète.

4.1. Environnement d'expérimentation

Nous avons effectué nos expériences dans un cluster de 10 machines que nous avons simulé avec la plateforme SimJava [18] dans un ordinateur de 250 Go de disque dur avec 2 Go de mémoire centrale et 2.5 GHz de processeur fonctionnant avec le système d'exploitation Windows XP. Nous avons instancié le service de requêtes complexes avec java et jdbc pour gérer les communications entre les différents participants et l'accès aux bases de données locales. Comme système de gestion de base de données, dans chaque participant, nous avons utilisé H2 [15]. H2 est un SGBD en mémoire principale. Ce qui permet d'accélérer les traitements sur la base locale. Nous avons travaillé avec une base de données miroir (dump) du GBIF que nous avons téléchargé. Nous avons réparti les données de sorte que chaque base contient en moyenne 40000 enregistrements d'occurrences d'espèces. On suppose que la taille maximale d'une base est de 100000 enregistrements. Ce paramétrage concerne seulement les expériences

avec notre solution. Pour les expériences avec la réplication totale, chaque base dispose de 500000 enregistrements.

Nous avons mesuré les temps de réponses de quelques requêtes décomposables et non-décomposables dans notre solution, que nous avons comparée avec les temps de réponse des mêmes requêtes pour une solution de réplication complète.

4.2. Expériences

Nous avons mesuré le temps de réponse de chaque requête. Les requêtes testées sont composées d'opérations de sélection et de jointure (pour lier le nom de l'espèce à ses occurrences). Dans notre solution, le temps de réponse d'une requête décomposable correspond à la valeur de temps la plus fréquente pour dix exécutions successives de la même requête. Pour les requêtes non-décomposables, on a pris la moyenne des temps de réponse de deux exécutions de la même requête. En effet, la première exécution nécessite un transfert des données en un seul endroit et est plus longue que la deuxième pour laquelle toutes les données impliquées sont déjà en un seul participant. Concernant l'éventuelle solution de réplication complète sur tous les participants, on a chargé sur la base locale, toutes les données dont nous disposons. Donc pour cette solution de réplication totale, tous les fragments requis pour le traitement d'une requête sont disponibles dans le participant local. On considère le temps de réponse d'une requête comme étant le délai entre l'instant de soumission de la requête et l'instant où l'utilisateur commence à lire ses résultats.

4.3. Résultats et discussions

4.3.1. Requêtes décomposables

Requête	Schéma de répartition des fragments dans notre solution (P1=Site local)	Notre solution (ms)	Réplication complète (ms)
R (A)	A : P2	735	5813
R (B)	B : P1	422	5828
R (A, B)	A : P2 ; B : P1	922	6078
R (A, C)	A : P2 ; C : P3	1000	6047
R (A, B, C, D)	A : P1 ; B : P2 ; C, D : P3	1203	6109

L'obtention des résultats escomptés pour toutes les requêtes et ceci dans des délais raisonnables avec un temps de réponse maximal de 1203 ms montre la faisabilité de

notre solution. Pour chaque requête évaluée, on voit nettement que le temps de réponse obtenu avec notre solution est toujours meilleur que celui obtenu avec la solution de réplication complète. La performance de notre solution est au moins 5 fois (6109/1203) plus rapide que la solution de réplication complète pour l'exécution des requêtes décomposables. Ceci est dû au fait que les tailles des relations traitées sont plus petites avec notre solution qu'avec la réplication complète. Cela est d'autant plus vrai pour les requêtes qui nécessitent des opérations de jointure. De plus, dans le cas d'une requête décomposable, les résultats sont retournés à l'utilisateur dès qu'un premier participant a fini son traitement. Ce qui n'est pas possible avec la réplication complète où il faut attendre la fin du traitement global de la requête avant de répondre à l'utilisateur.

4.3.2. Requêtes non- décomposables

Requête de décompte du nombre d'occurrence	Schéma de répartition des fragments dans notre solution (P1=Site local)	Notre solution (ms)	Réplication complète (ms)
R (A)	A : P1	781	6922
R (B)	B : P2	1000	6735
R (A, B, C)	A, B, C : P1	1031	6891
R (A, B, C)	A, B, C : P2	1079	6828
R (A, B, C)	A : P1 ; B, C : P2	1232,	6812
R (A, B, C)	A, B : P2 ; C : P3	1329	6844
R (A, B, C, D)	A, B : P2 ; C : P1 ; D : P3	1797	6906

Ce tableau des temps de réponse montre que les résultats de notre solution sont encore meilleurs pour chaque requête non-décomposable donnée. En effet, aucun temps de réponse n'a atteint les 2 secondes avec notre solution alors que dans le cas de la réplication complète tous les temps dépassent les 6.5 secondes. Là, il importe aussi de noter que nous n'avons pas effectué la réplication de toutes les occurrences du GBIF (400 millions) mais nous avons travaillé avec une portion de 500 mille occurrences. Ce qui montre que si on avait travaillé avec la base entière du GBIF, on aurait des temps encore beaucoup plus élevés pour la solution de réplication complète. D'où l'importance d'adopter une réplication partielle.

Les résultats obtenus par notre expérimentation, tant pour les requêtes décomposables que pour les requêtes non-décomposables prouvent l'efficacité de notre solution pour sa faisabilité et pour ses performances. Ces expériences ont permis de traiter des requêtes

jusqu'ici non supportées par le portail du GBIF. Ce qui permet d'enlever la limite de l'expressivité des requêtes du GBIF actuel avec des temps de réponse acceptables. En outre notre proposition contribue à l'amélioration des performances avec la répartition des accès et la parallélisation des traitements qui réduit de façon considérable les problèmes liés à la congestion et la charge de travail du système central du GBIF.

5. Related work

Beaucoup de travaux ont concerné le traitement parallèle de requête dans les environnements distribués. En 2000 **Kossmann et al** [22] ont fait un état de l'art du domaine en 2000 qui récapitulait une bonne partie des travaux et techniques qui existaient jusqu'à cette époque. La plupart de ces techniques sont utilisés dans des technologies récentes pour améliorer l'optimisation du traitement et/ou de l'utilisation de la bande passante. Parmi ces techniques, on peut citer le *row blocking* pour optimiser l'utilisation de la bande passante lors du transfert de données, le *data shipping* ou *query shipping* pour le transfert respectif de données et de traitement. En autres techniques étudiées par Kossmann et al, on peut ajouter la réplication et le caching. Dans notre approche, nous avons mis en place des mécanismes combinant *row blocking* et *caching* et *data shipping* pour minimiser le coût de transfert de données ainsi le *query shipping* pour exploiter au mieux le parallélisme dans notre contexte. Des travaux considérables [6][7][8][20][21] ont continué d'être menés pour aboutir à HadoopDB [13] en 2010 en passant GFS, Bigtable, HDFS, MapReduce, Hadoop, Hive, etc. Si GFS [20] et HDFS [21] sont des systèmes de gestion de fichiers adaptés à des environnements distribués, MapReduce [8] est un mécanisme de traitement parallèle de requête qui consiste à décomposer la requête en un ensemble de tâches réparties à travers les sites disposant des données concernés. Beaucoup de systèmes inspirés des principes de MapReduce ont ainsi vu le jour comme Hadoop [6]. Hadoop est une implémentation de MapReduce qui utilise le système de fichiers HDFS et offre de bonne performance pour le traitement parallèle sur de gros volume de données (plusieurs To). Cependant il n'est pas efficace pour des traitements locaux sur de petites quantités de données (quelques Mo). Hive [7][11] et HadoopDB [13] qui sont des dérivés de Hadoop, ont comme lui, les caractéristiques d'être performants avec les gros volumes de données et médiocres avec les petites quantités de données. Hive offre une interface proche du langage SQL pour une interrogation plus aisée des données, alors HadoopDB stocke ses données dans des bases de données relationnelles avec une implémentation avec PostgreSql pour bénéficier des performances des systèmes relationnels. Il utilise Hadoop et Hive pour la gestion des communications afin de bénéficier du parallélisme. Hadoop et Hive sont fortement utilisés par Yahoo et Facebook pour la gestion de leurs quantités énormes de données (plusieurs centaines de terabytes) [11]. Ses systèmes disposent d'une architecture Master/Slave où toutes les requêtes passent par le nœud maître appelé *Namenode*, qui stocke métadonnées sur le système global avant d'être traitées par les nœuds esclaves qui contiennent les données. La plupart de ces solutions prônent une

centralisation de la coordination de l'exécution des requêtes. Cette architecture est différente de celle de notre solution où chaque nœud est à la fois maître et esclave où les métadonnées sont stockées dans un catalogue synchronisé et les données dans les bases de données locales. Ces systèmes diffèrent aussi de notre solution du fait qu'ils ne donnent de bonnes performances pour le traitement de requêtes sur une petite quantité de données alors notre solution a pour objectif de traiter toute requête quelque soit sa complexité et le volume de données impliquées dans des délais raisonnables. La plupart de ces solutions prônent une centralisation de la coordination de l'exécution des requêtes.

6. Conclusion

Ce travail présente une solution de décentralisation du portail GBIF dans l'optique d'améliorer les performances liées au traitement des requêtes et de permettre aux utilisateurs de définir de nouveaux besoins sur les données. Nous avons proposé une distribution des données sur plusieurs participants qui collaborent pour améliorer les performances du système. Nous avons défini une architecture décentralisée pour traiter efficacement les requêtes des utilisateurs. Elle présente l'avantage de fonctionner sans modifier le portail GBIF existant. Nous avons proposé une stratégie de fragmentation et réplique dynamique des données et des mécanismes de traitement de requêtes dans ce contexte. Nous avons effectué les premières expérimentations de notre solution. Les résultats obtenus montrent la faisabilité de notre solution et son efficacité pour quelques requêtes typiques de l'usage en biodiversité. L'objectif principal de ces premières expériences, est de montrer la faisabilité de notre solution.

L'étude des performances notamment dans un environnement à large échelle comme le cloud constitue l'objet dans la prochaine phase de nos travaux. Pour cela, nous allons d'abord approfondir nos recherches sur les limites et les avantages de la gestion des données [10][11] dans le cloud ensuite améliorer nos mécanismes de traitement de requêtes avec une description plus formelle et des expérimentations plus complètes dans une infrastructure où les nombres de machines et de requêtes ainsi des quantités de données seront variés. L'amélioration de ces performances permettra d'adopter une solution dynamique pour passer à l'échelle avec l'augmentation des données intégrées et la croissance des usagers du GBIF.

7. Bibliographie

- [1] D. Hobern, *GBIF Biodiversity Data Architecture, GBIF Data Access and Database Interoperability (DADI)*, 2003.
- [2] GBIF, *GBIF Annual Report 2009*, pages 25-42, 2010.
- [3] GBIF international www.gbif.org, data.gbif.org et GBIF France, www.gbif.fr.

- [4] H. Saarenma. *Sharing and Accessing Biodiversity Data Globally through GBIF*, *ESRI User Conf.*, 2005.
- [5] *The GBIF Data Portal: A practical “hands-on”* accessible au <http://data.gbif.org>.
- [6] Apache Hadoop <http://wiki.apache.org/hadoop>.
- [7] Hive wiki <http://wiki.apache.org/hadoop/hive>.
- [8] J. Dean, S. Ghemawat. *MapReduce: simplified data processing on large clusters*, *CACM*, 2008.
- [9] M. Brantner et al. Building a Database on S3, *SIGMOD*, 2008.
- [10] D. J. Abadi: *Data Management in the Cloud: Limitations and Opportunities*. *IEEE Data Eng. Bull.* 32(1), 2009.
- [11] A. Thusoo et al. Hive - A Petabyte Scale Data Warehouse Using Hadoop, *ICDE*, 2010.
- [12] K. Bajda-Pawlikowski et al., *Efficient processing of data warehousing queries in a split execution environment*. *SIGMOD*, 2011:
- [13] A. Abouzied et al. *HadoopDB in action: building real world applications*. *SIGMOD*, 2010.
- [14] N. Bame. *Traitement de requêtes pour les données du GBIF : Utiliser un Cloud pour améliorer les performances des requêtes*, Mémoire de DEA d’Informatique, UCAD, 2011.
- [15] H2 Database Engine <http://www.h2database.com>.
- [16] D. Agrawal, S. Das, A. El Abbadi: Big data and cloud computing: current state and future opportunities. *EDBT* 2011.
- [17] N. Bame, H. Naacke, I. Sarr, S Ndiaye, *Architecture répartie à large échelle pour le traitement parallèle de requête de biodiversité*, 11th African Conference on Research in Computer Science and Applied Mathematics (CARI’12), Algiers, Ageria, pages 143-150, 2012.
- [18] www.simjava.com
- [19] T. Loukopoulos and I. Ahmad, *Static and adaptive data replication algorithms for fast information access in large distributed systems*, *ICDCS*, pp 385-392, 2000
- [20] S. Ghemawat et al., *The Google file system*, *SIGOPS Oper. Syst. Rev.* 37, 2003.
- [21] D. Borthakur, *The Hadoop Distributed File System: Architecture and Design*, The Apache Software Foundation, 2007
- [22] D. Kossmann et al., *The State of the Art in Distributed Query Processing*, *ACM Computing Surveys*, 2000, pp. 422–469.

Posters

Reconstruction volumique d'images dans un milieu poral avec java 3D

LACGAA Chérif Ndiaye LACGAA Ndièye Fatou Ngom et EPT Oumar Niang

Contexte et Objectifs

L'espace poral est le vide qui existe dans un milieu poral.
2 sortes d'images: images 2D et images 3D.



Fig1: image 3D et 2D

- Fournir une application capable de reconstruire une image 3D à l'aide de points 2D autorisant l'analyse de toute portion 2D ou 3D de l'image initiale
- Possibilité de visualiser l'espace poral

Méthodologie

- Superposition d'images 2D de sol permettant de visualiser l'espace poral par seuillage.
- graphe de scène

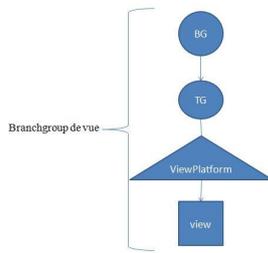
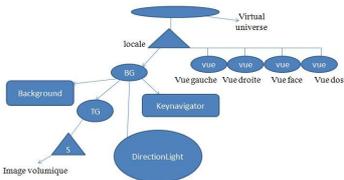


Fig2: graphe de scène de l'application

Résultats

- fig3: reconstruction de points aléatoires.

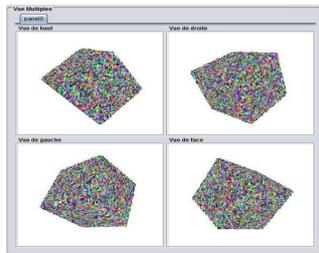
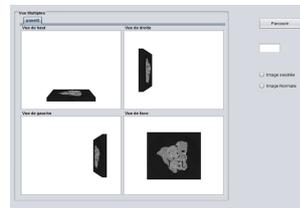


Fig4 et5: reconstruction volumique d'une image 3D suivant 4 plans reconstruite à partir d'images 2D de sol. (fig4) et visualisation de l'espace poral (fig5)



• fig6: images utilisées : images tomographiques de sol

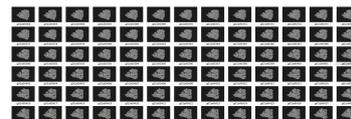
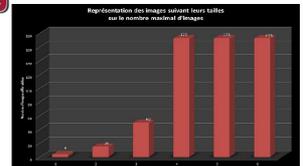


Fig7: Histogramme: Nombre d'images affichables en fonction de leur taille sur une machine avec RAM (4Go):



Conclusions

- application de reconstruction volumique d'images de milieu poral.
- Utilisation de java 3D

Perspectives

- Analyse spectrale de l'image 3D (ensemble de voxels).
- Extraction des points de contours afin de modéliser l'espace poral 3D pour simuler la dynamique biologique ou morphologique du sol dans le cas d'image statique et dans le cas d'images dynamiques.

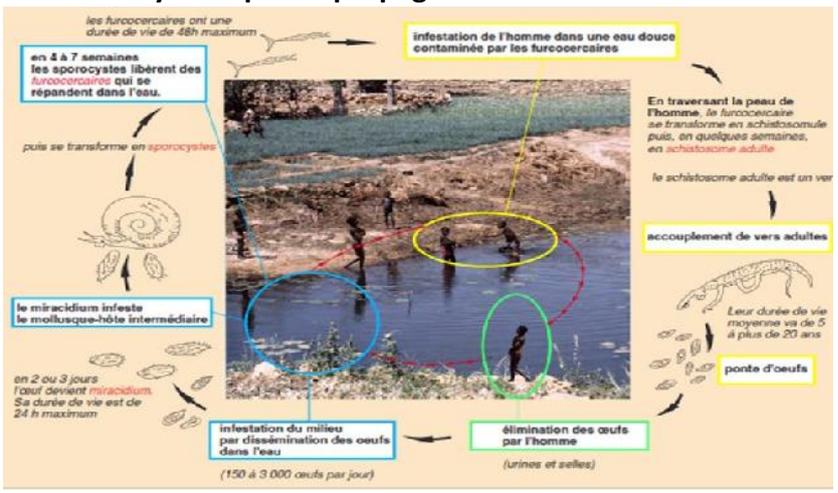
Références

- Ndiaye C.H.T.C., Reconstruction volumique d'images de sol avec l'api java3D de java, Master, UCAD, 2012
- Ngom , N.F., Monga, O., Mahmoud, MOM, Garnier, P, 3D shape extraction segmentation and representation of soil microstructures using generalized cylinders. J. computers and geosciences, 39. p.50-63

Modélisation à base d'agents de la propagation de la bilharziose

Problématique: Bilharziose: maladie parasitaire retrouvée en zones tropicales

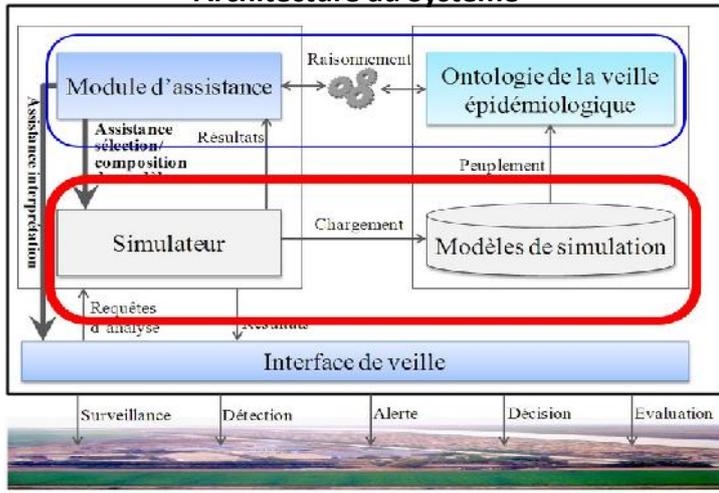
Dynamique de propagation de la bilharziose



Contexte:

- Etude de phénomènes épidémiologiques
- Mise en place d'un Système de veille épidémiologique pour la bilharziose

Architecture du Système



Objectifs :

- revisiter, compléter et adapter au formalisme « agent », les modèles (principalement mathématiques) de dynamiques intervenant dans la propagation de la Bilharziose
- faciliter l'interopérabilité et la composition de ces modèles de simulation de dynamiques
- assurer la distribution/répartition de leurs simulations dans un environnement distribué

Architecture à deux niveaux

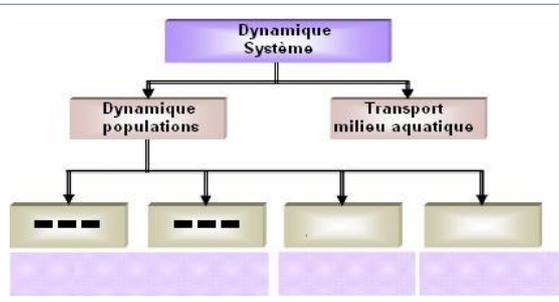
- Haut niveau:** constitué des approches qualitatives, il permet de s'abstraire de certaines propriétés détaillées du phénomène surveillé
- Bas niveau:** intégrant les approches numériques, décrit substantiellement les caractéristiques physiques et dynamiques du phénomène.

Méthodologie:



Un premier modèle à base d'agents du phénomène à partir de modèles mathématiques

Etape 1



Etude séparée des différents dynamiques intervenant au phénomène

Etape 2

Finalisation du modèle à base d'agents avec une étude de composition des sous modèles élaborés.

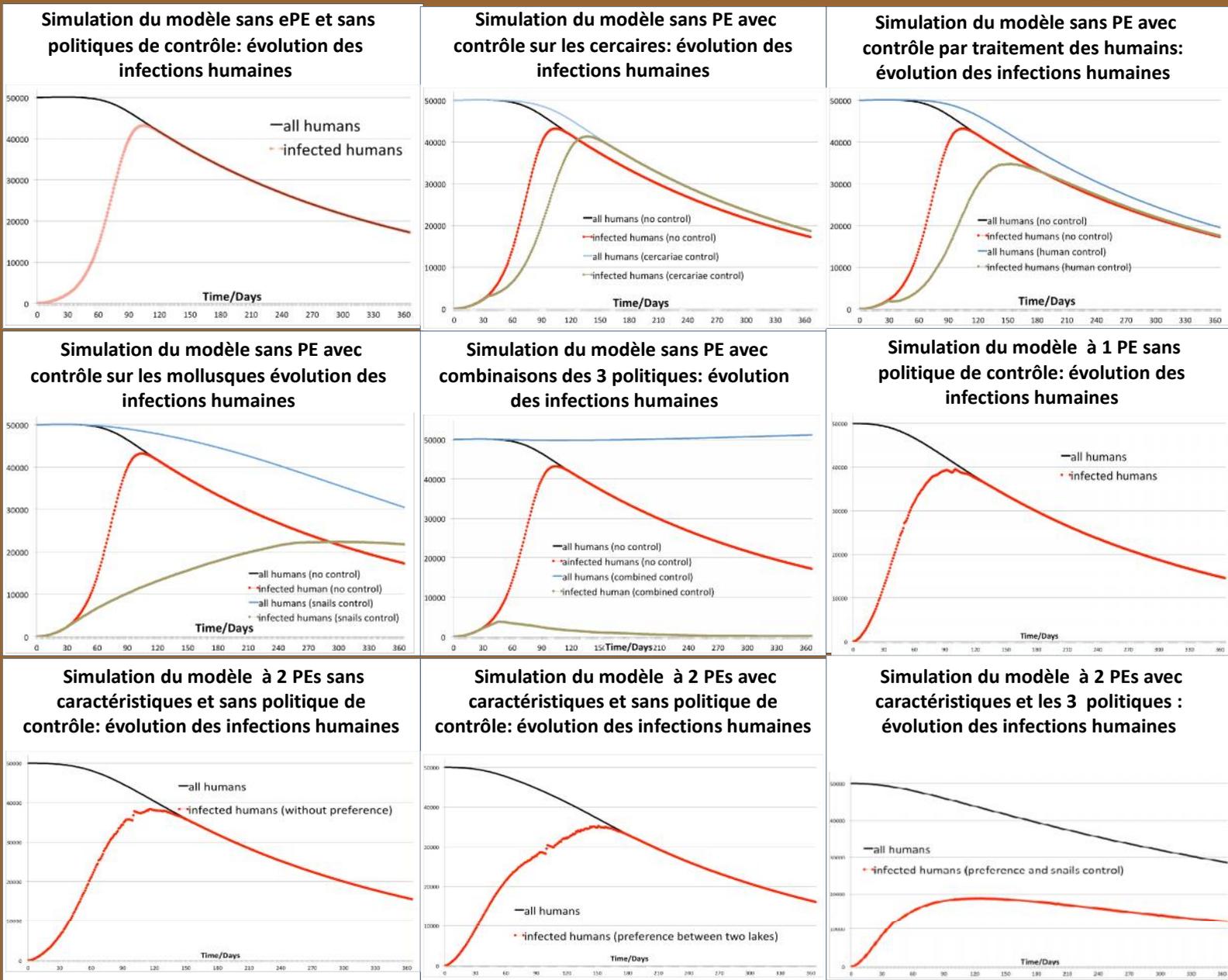
Etape 3

Etude et mise en place de l'architecture finale d'agents pour une simulation dans un environnement distribué

Etape 4

Premiers Résultats:

Ce qui suit concerne l'étape 1 de nos travaux. L'objectif était d'évaluer et de quantifier l'impact de la prise en compte de l'environnement (composé d'habitations et de points d'eau) sur la modélisation des politiques de contrôles de la maladie. D'abord, nous avons établi un premier modèle à base d'agents ne prenant pas en compte l'environnement. Ensuite, avec ce modèle, nous établissons une première simulation pour faire une analyse de sensibilité de certaines stratégies de lutte sur la propagation de la bilharziose. Puis, nous élargissons notre modèle en intégrant un environnement avec un seul point d'eau (PE). Par la suite, ce modèle est élargi en un environnement avec deux points d'eau sans caractéristiques distincts. Enfin, avec ce dernier modèle, nous faisons doter aux points d'eau des caractéristiques leurs permettant d'attirer certaines populations.



Conclusion:

- La prise en compte de l'environnement est importante dans le processus de modélisation de la bilharziose
- Les politiques de contrôle doivent être définies selon les caractéristiques environnementales de chaque système

Perspectives:

- étude séparée et modélisation des différentes dynamiques pouvant intervenir dans la propagation de la Bilharziose (dynamique des populations, transport en milieu aquatique, ...)
- Proposition d'une architecture de composition des différents sous modèles
- Déploiement en environnement distribué.

Démonstrations/Tutoriels

Vision artificielle et Traitement d'Images (VATI)
Bibliothèques de manipulation d'images avec Java

Ndèye Fatou NGOM
Laboratoire LACGAA
UCAD

1 Objectifs

Le VITA est devenu un domaine qui a des tentacules dans presque toutes les autres disciplines liées aux technologies numériques. Cet essor a été impulsé par des champs d'applications aussi divers que la médecine, la visualisation scientifique, la réalité virtuelle, les télécommunications et les effets spéciaux.

L'objet de ce tutoriel est d'introduire à quelques bibliothèques de manipulation d'images avec Java en 2D et en 3D.

2 Qu'est ce qu'une image

Une image est un ensemble structuré d'informations sur un support qui a une signification pour l'oeil. C'est une représentation visuelle d'une scène qui peut être constituée d'objets du monde réel (personne, animal, microorganismes, parties du corps humain, images météo...) ou non (oeuvres d'arts, images de synthèse..).

- Elle peut évoluer en fonction d'un paramètre (temps en général), on parle alors d'**images dynamiques**.
- Elle peut aussi ne pas changer, on parle alors d'**images statique**.
- L'impression d'**image 3D** est obtenue en ajoutant une profondeur.

3 Les branches du VATI

- **Vision par ordinateur (computer vision)** : conception et l'analyse des aspects calculatoires et algorithmique des processus d'acquisition, de traitement et d'interprétation des images numériques.
- **Traitement d'image (image processing)** : art d'effectuer des transformations sur une image numérique en vue d'en améliorer la qualité ou d'en extraire des informations pour des applications spécifiques.
- **Vision Artificielle** basée sur la réalité virtuelle qui est l'ensemble de technologies et de procédés qui permettent à un concepteur de créer un environnement entièrement généré par ordinateur.
- **Infographie (computer graphics)** : art de générer des images à partir d'un modèle géométrique qui prend en compte la forme, la texture et les couleurs des objets dans une scène réelle ou non.

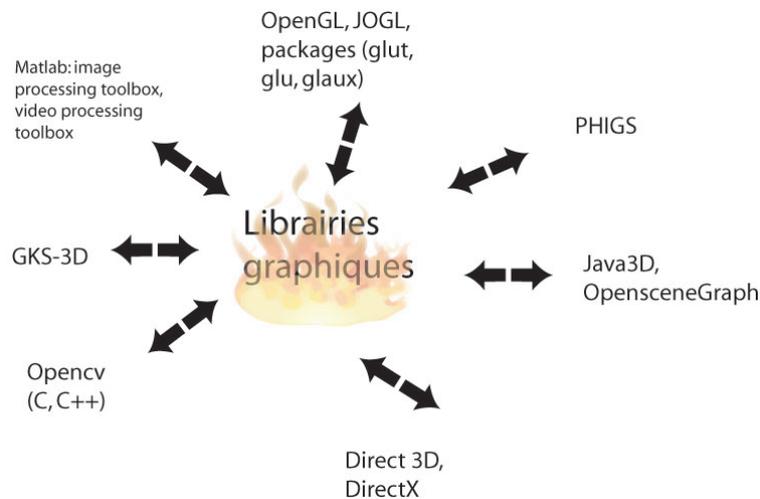


FIGURE 1 – Quelques exemples de librairies graphiques

4 Bibliothèques avancées de manipulation d'images avec Java

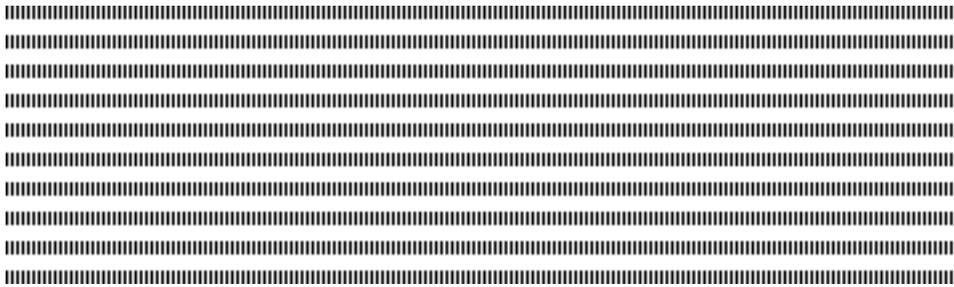
Les bibliothèques de manipulations de manipulations d'images avec Java peuvent être classées en trois catégories : les bibliothèques natives, les bibliothèques 2D et les bibliothèques 3D. Après avoir présenté un exemple de squelette dans chaque cas, nous présenterons une illustration avec trois applications

1. Segmentation interactive d'images 2D fonction d'un paramètre dynamique (Ndiaye et al, 2012),
2. Détection interactive de formes (humaines ou géométriques) dans une image fixe ou dans une vidéo choisie par l'utilisateur (Ouedraogo et al, 2013),
3. Reconstruction volumique et génération de scènes 3D.

5 Prérequis

Pour suivre sans difficulté ce tutoriel, il est souhaitable de posséder des notions en POO et d'installer

- la bibliothèque **javacv** qui peut être téléchargée gratuitement à l'adresse suivante <http://code.google.com/p/javacv/>
- la bibliothèque **java3d** qui peut être gratuitement téléchargée à l'adresse suivante <http://java3d.java.net/binary-builds.html>
- Configurer l'EDI (de préférence NetBeans) pour l'utilisation des bibliothèques.



Introduction à la programmation des capteurs sans fil

Traore Mahamadou* — Pottier Bernard** — Thiaré Ousmane*

* Département Informatique
Université Gaston Berger
Ville Saint-Louis
Pays Sénégal
{mahamadou.traore,ousmane.thiare}@ugb.edu.sn

** Département Informatique
Université de Bretagne Occidentale
Ville Brest
Pays France
bernard.pottier@univ-brest.fr



RÉSUMÉ. Les réseaux de capteurs sans fil (RCSFs) sont aujourd'hui largement utilisés dans la collecte et la transmission de données dans plusieurs domaines: militaire, agriculture, médecine, domotique, environnement etc. Ils sont constitués d'un ensemble de nœuds capable de percevoir l'information, de la traiter puis de la transmettre et ceci de façon coopérative. Nous proposons dans ce tutoriel comment programmer et déployer des réseaux de capteurs sans fils avec des plateformes de prototypage open source.

ABSTRACT. The wireless sensor networks (WSNs) are now widely used in the collection and transmission of data in several areas: military, agriculture, medicine, automation, environment etc. They consist of a set of nodes that can receive information, process it and then transmit it in a cooperative way. We propose in this tutorial how to program and deploy wireless sensor networks with open source electronics prototyping platform.

MOTS-CLÉS : RCSFs, ZigBee, XBee, 802.15.4

KEYWORDS : WSNs, ZigBee, XBee, 802.150.4



1. Introduction

Les réseaux de capteurs sans fil (RCSFs) sont aujourd'hui largement utilisés dans la collecte et la transmission de données dans plusieurs domaines : militaire, agriculture, industriel, médecine, domotique, environnement etc. Ils sont constitués d'un ensemble de nœuds capable de percevoir l'information, de la traiter puis de la transmettre et ceci de façon coopérative[1].

Une bonne conception des réseaux de capteurs requiert une parfaite maîtrise de :

- 1) la mesure physique, la validation et le calibrage des capteurs ;
- 2) la prise en main des micro systèmes (MCU) qui pilotent les acquisitions, pré-treatent les informations et préparent leur transmission ;
- 3) l'apprentissage des communications radio avec un grain de détails pratiques assez grossier, et les fondamentaux de la communication radio (802.15.1)[2] ;
- 4) la représentation des implantations de capteurs dans l'environnement, et des modèles environnementaux à observer et contrôler.

Une expérience a été effectuée à Brest sur 4 jours fin Octobre 2012, les documents et plans de cours restent disponibles en ligne[3] par login anonyme. Cette école a été une occasion pour des scientifiques, de réfléchir sur des thématiques d'instrumentations et de modélisations avec les RCSFs.

Nous proposons dans ce tutoriel, comment programmer et déployer un RSCF avec des plateformes de prototypage open source Arduino[4] et/ou Libelium[5] peu onéreuses et des modules de transmission radio XBee[6].

- Utiliser les plateformes de prototypage électronique open source pour déployer des infrastructures de collecte de données. ;
- Créer un réseau ZigBee avec des modules de transmissions radio XBee ;
- Se familiariser avec le mode API XBee pour créer des RCSFs ;
- Configurer des passerelles qui interconnectent les différents réseaux.

2. Moyens matériels et logiciels

La majeure partie des planches et travaux pratiques de l'école restant disponible, nous proposons de les mettre à disposition d'auditeurs, en mettant en place un tutoriel *pratique*. Dans ce tutoriel nous montrons ceci :

- 1) les assemblages de capteurs bon marché incluant des MCU Atmel, et des capteurs simples : température, hygrométrie, infra rouge, gaz etc. ;
- 2) les outils de programmation ;
- 3) les communications radio sur des supports "XBee" 802.15 pair a pair ;
- 4) le maquettage d'une distribution de capteurs sur carte.

Les outils matériels et logiciels utilisés pour cela sont :

- des plateformes open source de prototypage Arduino et Libelium ;
- des capteurs de température, d'humidité et de gaz ;
- des modules XBee pour une transmission radio des données collectées ;
- un environnement de développement des plateformes Arduino, Libelium.

Ces travaux pratiques pourraient être accompagnés d'ouvertures recherche, par exemple implantation de machine virtuelle Occam (TVM) sur les MCU, par exemple "sniffers" de paquets sur le réseau.

La mise en place requiert la présence d'au moins 2 personnes, une salle d'au moins 10 PC disposant d'une partition Linux dont nous avons le contenu.

3. conclusion

Dans ce tutoriel, nous proposons une série de manipulations réalisées avec des plateformes de prototypage propriétaires, peu onéreuses, des modules XBee pour une transmission radio des données collectées avec des capteurs. Des infrastructures de collecte de température, de son ou de niveau de gaz peuvent ainsi être déployées.

Ce tutoriel permettra aux étudiants de se familiariser avec les outils matériels et logiciels pour programmer des capteurs et RCSFs.

4. Bibliographie

- [1] YICK, JENNIFER, MUKHERJEE, BISWANATH , GHOSAL, DIPAK, « Wireless sensor network survey », *Comput. Netw.*, vol. 52, n° 12, 2008.
- [2] STANDARD 802.15.4, « Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPANs) ».
- [3] [HTTP://ECOLE-CAPTEURS.UNIV-BREST.FR](http://ecole-capteurs.univ-brest.fr), « Ecole internationale "Réseaux de Capteurs, Mesures et Environnement" ».
- [4] [HTTP://WWW.ARDUINO.CC](http://www.arduino.cc), « Arduino Website ».
- [5] [HTTP://WWW.LIBELIUM.COM](http://www.libelium.com), « Libelium Website ».
- [6] [HTTP://WWW.DIGI.COM](http://www.digi.com), « Digi Website ».

Index des auteurs

Abdessalem, Talel	70
Ba, Mandicou,	86
Bame, Ndiouma	126
Ben Meriem, Tayeb	110
Bendraou, Reda	102
Blanc, Xavier	102
Camara, Gaoussou	30
Cambier, Christophe	139
Chaudet, Claude	110
Cisse, Papa Alioune	118, 139
Cretton, Fabian	22
Declercq, David	94
Dembele, Jean Marie	118, 139
Despres, Sylvie	30
Diallo, Gayo	54
Diané, Ibrahima	62
Dieng, Ndeye Amy	110
Diongue, Dame	14
Diop, Ibrahima	118
Diouf, Madiagne	94
Djedidi, Rim	30
Dramé, Khadim	54
Fall, Ibrahima	102
Farssi, Sidi Mohamed	46
Flauzac, Olivier	38, 86
Gervais, Marie-Pierre	102
Gueye, Bassirou	38
Gueye, Modou	70

Kacimi, Rahim	62
Le Calvé, Anne	22
Liu, Zhan	22
Lo, Moussa	30, 118, 139
Makhloufi, Rafik	86
Mammeri, Zoubir	62
Mendy, Gervais	46
Mougin, Fleur	54
Naacke, Hubert	70, 126
Ndiaye, Abdourahmane	46
Ndiaye, Chérif	138
Ndiaye, Samba	126
Ngom, Ndeye Fatou	138, 142
Niang, Ibrahima	38, 62, 86
Niang, Oumar	138
Nolot, Florent	86
Ouya, Samuel	46, 94
Pottier, Bernard	144
Rabat, Cyril	38
S	
Sarr, Cheikh	6
Sarr, Idrissa	126
Thiaré, Ousmane	14, 144
Toutain, Laurent	110
Traore, Mahamadou	144
Wade, Ahmed Mouhamadou	78