

Méthodes Formelles

1 Automates sur les mots de mots

Responsable : Frédérique Carrère et Géraud Sénizergues

lieu : LaBRI

téléphone : 05.40.00.84.25 ; 05.40.00.35.26

e-mail : carrere@labri.fr ges@labri.fr

équipe : Méthodes Formelles, thème : Logiques, Graphes et Langages

Mots-clés : Automates finis, Mots de mots

Description du sujet

Contexte : Un *mot de mots* est une suite finie de symboles pris dans un alphabet infini X_2 qui est lui-même formé de tous les mots ordinaires sur un alphabet fini X (notons $X^{[2]}$ l'ensemble de ces objets). Un automate à pile de piles est un automate dont la mémoire est constituée de mots de mots et dont les transitions n'effectuent que des opérations restreintes sur ces mots : empilement ou dépilement d'ordre 1 ou 2 ([Mas76]). L'étude de ces automates a conduit à s'intéresser aux mots de mots pour eux-mêmes et à définir des expressions rationnelles ([Car05]) et des automates finis ([Fra05]) sur $X^{[2]}$.

Le but de ce mémoire serait d'approfondir les différentes notions possibles d'automate fini sur $X^{[2]}$, puis sur les tuples d'éléments de $X^{[2]}$ puis d'en déduire une notion viable de *structure automatique* sur les mots de mots, en un sens analogue à celui employé par [Kho94] dans le cas des mots.

Objectifs :

1. Lire le chapitre II.6 de la thèse de Fratani et l'exposé [Sen-Sata08]. Comprendre la notion d'automate fini, avec oracle rationnel, sur le groupe libre et, par bijection, sur les mots de mots.
2. Établir une bijection entre les mots de mots et les arbres planaires étiquetés, et en déduire une nouvelle notion d'automate fini sur les arbres planaires ; on pourra s'inspirer notamment de [Sen-H.Comon08].
3. Faire l'inventaire des propriétés de clôture des langages de mots de mots (ou d'arbres planaires) reconnus par de tels automates.
4. Comparer le pouvoir d'expression des automates de Fratani (traduits sur les arbres) avec celui des automates finis classiques.
5. Éventuellement étendre la notion d'automate fini sur les mots de mots pour obtenir le même pouvoir d'expression que les automates d'arbres finis.

6. Examiner diverses notions de transductions rationnelles pour les mots de mots ; on pourra, par exemple, définir une notion adéquate de *superposition* de deux mots de mots et considérer les ensembles de couples dont la superposition est reconnue par un automate de mots de mots.
7. Étudier la possibilité de définir à partir d'un choix adéquat de la notion de superposition, une notion de *structure automatique de niveau 2*.
8. Une perspective à plus long terme serait d'étendre à tous les niveaux $k \geq 2$ la notion de structure automatique de niveau k puis de faire une investigation systématique de ces structures ; par exemple, tous les graphes de calcul d'automates à pile de niveau k devraient être automatiques de niveau k .

Références bibliographiques

- A.N. Maslov. Multilevel stack automata. *Problemy Peredachi Informatsii* 12, pages 38–43, 1976.
- Arnaud Carayol. Regular sets of higher-order pushdown stacks. In *Mathematical foundations of computer science 2005*, volume 3618 of *Lecture Notes in Comput. Sci.*, pages 168–179. Springer, Berlin, 2005.
- S. Fratani. *Automates à piles de piles ... de piles*. PhD thesis, Bordeaux 1 university, 2005. Chapter II.6.
- Bakhadyr Khossainov and Anil Nerode. Automatic presentations of structures. In *Logic and computational complexity (Indianapolis, IN, 1994)*, volume 960 of *Lecture Notes in Comput. Sci.*, pages 367–392. Springer, Berlin, 1995.
- G. Sénizergues. Higher-order pushdowns and monoids. *Talk given at SATA'08*, pages 1–46, 2008. Available at URL :<http://dept-info.labri.u-bordeaux.fr/~ges>.
- G. Sénizergues. Words and trees. *Talk given at H. Comon's reception of the CNRS Silver medal*, pages 1–66, 2008. Available at URL :<http://dept-info.labri.u-bordeaux.fr/~ges>.

2 Un outil de preuve sur les graphes

Responsable : Pierre Castéran

lieu : LaBRI

téléphone : 05.40.00.69.31

e-mail : pierre.casteran@labri.fr

équipe : Méthodes Formelles, thème : Logiques, graphes et langages, projet : Visidia

Mots-clés : Logique, preuves, graphes

Description du sujet

Contexte et objectifs : Ce travail se situe dans le cadre des projets Loco et Visidia (voir les liens ci-dessous). Le domaine d'intérêt de ces projets est l'étude d'une classe d'algorithmes distribués, présentés comme des systèmes de transformations de graphes étiquetés. Prouver la correction d'un tel algorithme est souvent une tâche complexe et possiblement entâchée d'erreurs, ce qui relativise la certitude que les algorithmes sont *vraiment* corrects.

Le projet Loco s'intéresse à la construction d'un ensemble d'outils pour faciliter la preuve *formelle* d'algorithmes distribués à base de réétiquetages de graphes. Ce projet est réalisé avec l'assistant de preuves *Coq*.

Le sujet de stage proposé s'attaque à un aspect important de ces preuves : Les preuves de correction des algorithmes étudiés font appel à des propriétés des graphes telles que l'acyclicité, la connexité, etc.

Nous considérons qu'une démonstration doit avoir plusieurs qualités essentielles :

- Être lisible et convaincante pour un lecteur (humain),
- Être suffisamment précise et complète pour être vérifiée par un programme informatique.

De plus, les outils d'aide à la démonstration devront faciliter – avec un degré raisonnable d'automatisation – l'obtention de ces preuves.

Le travail demandé comportera une prise de connaissance des recherches déjà effectuées dans cette direction (essentiellement les travaux de Duprat et Gonthier), et un choix de la représentation des graphes la plus adaptée à nos besoins.

Outre une argumentation scientifiquement étayée, on montrera la pertinence des décisions prises par l'écriture de lemmes simples sur les graphes étiquetés.

Ce projet s'adresse principalement aux personnes intéressées par l'interaction entre l'informatique fondamentale et les outils de vérification formelle. Les critères principaux de réussite sont la facilité de construction interactive de preuve ainsi que la lisibilité des textes de démonstration obtenus. Il faut donc avoir un intérêt à la fois pour le langage mathématique et les langages de programmation.

Références et liens utiles

- Page du projet Loco :
<http://www.labri.fr/perso/casteran/Loco/index.html>
- Page du projet Visidia :
<http://visidia.labri.fr>
- Site de l'assistant de preuves *Coq* :
<http://coq.inria.fr>
- Site du projet “Mathematical Components” :
<http://www.msr-inria.inria.fr/Projects/math-components>

- Page du livre sur *Coq* :
<http://www.labri.fr/perso/casteran/CoqArt/index.html>
- Documentation sur l’outil SSreflect :
<http://hal.inria.fr/inria-00258384/en/>
- La fameuse preuve du théorème des quatre couleurs :
<http://research.microsoft.com/en-us/um/people/gonthier/4colproof.pdf>
- Formalisation des graphes par Jean Duprat :
<ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR2001/RR2001-15.ps.Z>

3 Réécriture conservant la rationalité

Responsables : Irène Durand et Gérard Sénizergues

lieu : LaBRI

téléphone : 05.40.00.69.14 ; 05.40.00.35.26

e-mail : idurand@labri.fr ges@labri.fr

équipe : Méthodes Formelles, thème : Logique Graphes et Langages

Mots-clés : Réécriture, Automates

Description du sujet

Contexte : Un *système de réécriture* de mots *semi-Thuéien* [Klo92] est un ensemble S de règles $u_i \rightarrow v_i$ de remplacement d’un mot par un autre ; on dit qu’un mot w' *dérive* du mot w modulo S , lorsque w' peut être obtenu à partir de w par une suite finie de remplacements d’un facteur u_i par le mot correspondant v_i . Les systèmes semi-Thuéiens finis ont la puissance des machines de Turing ce qui entraîne que l’on ne peut pas décider, en toute généralité leurs propriétés algorithmiques telles que l’accessibilité, la terminaison ou la confluence.

Certaines sous-classes de systèmes semi-Thuéiens ont cependant des propriétés de *préservation* des familles de langages formels usuelles (les langages rationnels ou les langages algébriques) qui entraînent de meilleures propriétés de décidabilité. Ainsi, la sous-classe des systèmes *match-bounded*, introduite en [GHW04], préserve-t-elle la famille des langages rationnels et la sous-classe des systèmes *de bas-en-haut* [Dur-Sen07] inverse-préserve-t-elle la famille des langages rationnels.

Objectifs : Le but de ce mémoire est de comparer ces deux sous-classes et, si possible, de définir une sous-classe aussi naturelle que possible qui les *englobe* toutes les deux.

1. Lire la communication [GHW04] qui introduit la classe des systèmes *match-bounded*.
2. Lire la communication [Dur-Sen07] qui introduit la classe des systèmes *de bas-en-haut*.

3. Montrer, par des exemples, que ces classes de systèmes sont incomparables pour l'inclusion
4. Examiner si une version infinitaire, mais rationnelle, de la classe des systèmes match-bounded permettrait de contenir ou simuler la classe des systèmes de-bas-en-haut.
5. Étendre aux systèmes rationnels match-bounded le théorème de préservation de la rationalité.
6. Examiner la complexité de la transformation d'automates finis sous-jacente au théorème ci-dessus ; une utilisation de [Ben-Sak86] pourra être utile.
7. Implémenter, sous forme d'un module additionnel du logiciel Autowrite ([Dur02], <http://dept-info.labri.u-bordeaux.fr/~idurand/autowrite>) les algorithmes essentiels portant sur ces systèmes match-bounded étendus :
 - le semi-algorithme testant l'appartenance à cette classe
 - l'algorithme de transformation des rationnels
 - le semi-algorithme décidant la terminaison sur un rationnel de départ et produisant un "certificat", en cas de succès.
 Cette phase d'implémentation comprendra la réalisation d'un cahier des charges, de spécifications, d'une planification du travail et de tests. Le module sera écrit en Common Lisp pour une intégration naturelle dans Autowrite.
8. Une perspective à plus long terme serait d'étendre aux systèmes de réécriture de termes cette nouvelle classe de systèmes de réécriture de mots.

Autowrite

Références bibliographiques

- J.W. Klop. Term rewriting systems. In *Handbook of Logic in Computer Science, Vol. 2*, pages 1–116. Oxford University Press, 1992.
- A. Geser, D. Hofbauer, and J. Waldmann. Match-bounded string rewriting systems. *Journal Applicable Algebra in Engineering, Communication and Computing*, 15(3-4) :149–171, November, 2004.
- I. Durand and G. Sénizergues. Bottom-up rewriting is inverse recognizability preserving. In *Proceedings RTA'07*, volume 4533 of *LNCS*, pages 114–132. Springer-Verlag, 2007.
- M. Benois and J. Sakarovitch. On the complexity of some extended word problems defined by cancellation rules. *Inform. Process. Lett.*, 23(6) :281–287, 1986.
- I. Durand. Autowrite : A tool for checking properties of term rewriting systems (category : system description). In *Proceedings RTA'02*, volume 2378 of *LNCS*, pages 371–375. Springer-Verlag, 2002.

4 Calculs locaux et protocoles de population

Responsable : Mohamed Mosbah et Akka Zemmari

lieu : LaBRI

téléphone : 05.40.00.69.17

e-mail : mosbah@labri.fr

équipe : MF, projet : VISIDIA

Mots-clés : Algorithmique distribuée, réseaux de capteurs, graphes dynamiques

Description du sujet

Les calculs locaux, et plus particulièrement les systèmes de réétiquetage de graphes, présentent un modèle puissant pour l'étude, l'analyse et l'implémentation d'algorithmes distribués. On considère un réseau de processeurs ayant une topologie quelconque. Il est représenté par un graphe étiqueté, connexe, et non orienté, où les sommets correspondent aux processeurs et les arêtes aux canaux de communication. Les processeurs communiquent seulement par envoi de messages asynchrones. L'étiquette d'un sommet (resp. d'une arête) code l'état du processeur (resp. du lien de communication) correspondant. Un calcul sur un réseau, qui consiste donc à faire évoluer ces états, peut être décrit par des règles de réécriture de graphes. Chaque règle, définie par un graphe connexe et deux étiquetages de ce graphe, traduit une étape élémentaire du calcul effectuée sur une portion locale du graphe.

Un système de réécriture est défini par la donnée d'un ensemble dénombrable de règles. A partir de l'état initial d'un réseau, donc d'un étiquetage initial, l'exécution d'un algorithme distribué est modélisée par une suite d'applications de règles. Une comparaison ainsi qu'une hiérarchie entre types de règles a été effectuée dans l'équipe combinatoire et algorithmes. Nous nous intéressons dans le projet visidia à trois types de règles (LC0, LC1, LC2). L'implémentation automatique d'un algorithme décrit à l'aide de ces règles de réécriture amène, en particulier, à étudier les algorithmes probabilistes permettant de synchroniser deux sommets voisins ou bien un sommet avec l'ensemble de ses voisins et plus généralement un ensemble de sommets correspondant à une règle de réécriture. Ces procédures probabilistes ont été analysés et étudiés et sont implémentées dans la plate-forme logicielle visidia. Par ailleurs, nous avons étudié un modèle d'implémentation basé sur les agents mobiles. Toutefois, les graphes sou-jacents ont une topologie fixe. Nous nous intéressons dans ce projet à une adaptation de ce travail si la topologie du graphe n'est pas connue, et tout particulièrement dans le cas des protocoles de population (ou réseaux de capteurs). Il s'agit de définir des algorithmes basés uniquement sur l'interaction locale des noeuds qui peuvent être mobiles. Or cette notion est compatible

avec les calculs locaux. Nous proposons de s'appuyer sur nos résultats et nos techniques de calculs pour généraliser les travaux d'Angluin et al. En particulier, on souhaite étudier et analyser des modèles de déplacement probabilistes de capteurs pour avoir des résultats sur leurs synchronisations (ou leurs interactions locales). Ces interactions peuvent être réalisées de manière indépendante et se définissent donc par une loi de probabilité. Ceci nous permet de disposer de résultats généraux et de définir des classes d'algorithmes plus larges. Ce projet porte un aspect plus appliqué à travers l'implantation des algorithmes étudiés. Dans ce mémoire, on considère le cas de rendez-vous (communication uniquement entre deux noeuds voisins).

<http://visidia.labri.fr>

Méthodes Formelles - Combinatoire et Algorithmique

5 Étiquetage de connectivité dans les réseaux planaires 3-connexes avec pannes

Responsables : Bruno Courcelle et Cyril Gavoille

lieu : LaBRI

e-mail : courcell@labri.fr, gavoille@labri.fr

équipes : Méthodes Formelles / Combinatoire et Algorithmique

Mots-clés : graphes

Description du sujet

http://dept-info.labri.fr/~gavoille/Sujets-Stages-Theses/sujet_M2R-09.html

Combinatoire et Algorithmique

6 Graphes arbitrairement décomposables

Responsable : Olivier Baudon

lieu : LaBRI

téléphone : 05.40.00.69.21

e-mail : olivier.baudon@labri.fr

équipe : Combinatoire et Algorithmique, thème : Graphes et Applications

Mots-clés : Graphe, Décomposition arbitraire

Description du sujet

Si l'on modélise un réseau par un graphe, étudier le découpage d'un réseau en sous-réseaux revient à étudier les sous-graphes induits d'un graphe. Nous nous posons ici la question des réseaux que l'on peut décomposer en sous-réseaux de n'importe quelle tailles. Les seules propriétés que devront garantir les sous-graphes induits étant la connexité et une taille pré-définie.

Ce problème peut être formalisé de la façon suivante :

Soit k et n deux entiers strictement positifs tels que $n > k$. Une k -décomposition $\lambda = (\lambda_1, \dots, \lambda_k)$ de n est une suite de k entiers strictement positifs dont la somme vaut n . Si la valeur de k n'importe pas, on parlera simplement de *décomposition* de n .

Soit G un graphe. $G[S]$ dénote le sous-graphe induit de G par un sous-ensemble S de ses sommets.

Graphe λ -décomposable : Soit $G = (V, E)$ un graphe simple connexe à n sommets et $\lambda = (\lambda_1, \dots, \lambda_k)$ une k -décomposition de n . G est dit λ -décomposable si et seulement si il existe une partition de V en k sous-ensembles V_1, \dots, V_k tels que pour tout $i \leq k$

1. $|V_i| = \lambda_i$
2. $G[V_i]$ est connexe

Graphe arbitrairement décomposable : Soit $G = (V, E)$ un graphe simple connexe à n sommets. G est dit *arbitrairement décomposable* si et seulement si, pour toute décomposition λ de n , G est λ -décomposable.

Par exemple, toute chaîne et plus généralement tout graphe possédant une chaîne hamiltonienne est arbitrairement décomposable. A l'inverse, le graphe $K_{1,3}$ (trois sommets a, b, c reliés à un même sommet x) n'est pas arbitrairement décomposable. En effet, il n'admet pas de (2,2)-décomposition.

Ce problème a en particulier été étudié dans le cadre des arbres et un résultat important est que les arbres arbitrairement décomposables sont de degré maximum 4 [BF06].

Nous avons également étudié récemment une classe de graphes nommée k -ballons. Un k -ballon est un graphe composé de k chaînes reliées entre elles par leurs deux extrémités [BGW09].

Objectifs : Le but de ce stage recherche est d'étudier la valeur de k maximum pour qu'un k -ballon soit arbitrairement décomposable. Nous conjecturons que cette valeur est de 6. Une autre partie consistera à donner des valeurs de longueurs de chaînes pour lesquelles les k -ballons sont arbitrairement décomposables.

Références bibliographiques

BF06 D. Barth and H. Fournier. A degree bound on decomposable trees. *Discrete Applied Mathematics*, 306 :469–477, 2006.

BGW09 O. Baudon, F. Gilbert, and M. Woźniak. Recursively arbitrarily vertex-decomposable graphs. Technical Report 1452-08, LaBRI, Université Bordeaux 1, 2008. http://www.labri.fr/perso/lepine/Rapports_internes/RR-145208

7 Jeux combinatoires partisans de coloration de graphe

Responsable : Paul Dorbec, Éric Sopena

lieu : LaBRI

téléphone : 05.40.00.35.23 / 05.40.00.38.71

e-mail : paul.dorbec@labri.fr eric.sopena@labri.fr

équipe : Combinatoire et Algorithmique, thème : Graphes et Applications

Mots-clés : jeu combinatoire partisan, coloration de graphes

Description du sujet

Le sujet est décrit à l’adresse suivante.

http://www.labri.fr/perso/sopena/pmwiki/uploads/Proposals/2010_Dorbec_Sopena_M2R_Sujet_1.pdf

8 Codes identifiants adaptatifs dans les graphes

Responsable : Paul Dorbec

lieu : LaBRI

téléphone : 05.40.00.35.23

e-mail : paul.dorbec@labri.fr

équipe : Combalgo, thème : Graphes et Applications

Mots-clés : codes identifiants, adaptativité

Description du sujet

Dans une publication récente [1], Ben-Haim et al ont proposé une version adaptative du problème d’identification des sommets d’un graphe. Un code identifiant est un ensemble de sommets S tel que tous les sommets ont des voisinages différents dans S (non vide). On peut percevoir ce problème comme la recherche d’un ensemble de requête auprès de sommets du type “Existe-t-il un sommet défaillant dans ton voisinage ?” permettant d’identifier un sommet défaillant. Dans la version adaptative, on s’autorise à choisir les sommets ciblés par les requêtes après avoir reçu la réponse des requêtes précédentes. L’identification se décrit alors par un algorithme de requêtes qui s’adapte aux réponses plutôt que par un ensemble

de requêtes choisis initialement. Le nombre de requête du problème adaptatif est clairement majoré par la cardinalité d'un code identifiant.

L'objectif de ce stage sera d'identifier les résultats récents sur ce sujet dans la littérature, et de tenter de résoudre ce problème ou de proposer des bornes pour quelques familles de graphes simples, comme les arbres, les hypercubes, les produits de graphes... On pourra aussi tenter de caractériser les graphes pour lesquels cette adaptativité est très efficace (importante réduction du nombre de requêtes nécessaires) et ceux pour lesquels elle n'apporte aucun progrès par rapport aux codes identifiants.

Pour plus d'informations sur les codes identifiants et les codes adaptatifs, se référer aux documents mentionnés ci-après.

Références bibliographiques

- Y. Ben Haim, S. Gravier, A. Lobstein, J. Moncel, Adaptive identification in graphs, J. Comb. Th. A 115, 1114-1126 (2008)
- G. Cohen, I. Honkala, S.Litsyn, A. Lobstein, Covering codes, Elsevier, 1997

9 Modélisation des communications dans les plates-formes à grande échelle

Responsable : Lionel Eyraud-Dubois, Olivier Beaumont, Nicolas Bonichon

lieu : LaBRI

téléphone : 05.40.00.35.24

e-mail : lionel.eyraud-dubois@labri.fr, olivier.beaumont@labri.fr,
nicolas.bonichon@labri.fr

équipe : CombAlgo, thème : Algorithmique Distribuée, projet : INRIA CÉ-
PAGE

Mots-clés : Plates-formes à grande échelle, algorithmes d'approximation, complexité, modélisation, streaming

Description du sujet

Les plates-formes de calcul grande échelle (comme par exemple les systèmes de calcul volontaire comme BOINC [6]) constituent des solutions de plus en plus répandues pour effectuer des calculs de grande envergure. La conception d'algorithmes de gestion des ressources et des communications est un point clef pour améliorer l'utilisation de ces plates-formes. Pour cela, il est nécessaire d'avoir de bons modèles pour décrire le réseau d'une telle plate-forme, qui est souvent trop gros et trop complexe pour les modèles classiques.

Plusieurs travaux ont déjà été effectués dans ce sens, fondés sur des plongements géométriques [3,4,5] ou sur des modèles à base de graphes [1,2]. Ces tra-

vaux modélisent correctement la bande-passante et la latence, mais peu d'entre eux prennent en compte la contention (conflit d'utilisation de certains liens du réseau). Elle est pourtant un des facteurs principaux pour l'ordonnement des communications.

Le but de ce stage est d'explorer le modèle arborescent de Sequoia [4] en y ajoutant la prise en compte de la contention. Deux orientations sont envisagées :

- la conception d'algorithmes garantis pour le problème du broadcast en régime permanent [7] ;
- une validation expérimentale de ce modèle sur une plate-forme à grande échelle (comme PlanetLab par exemple).

Références bibliographiques

- 1 Lionel Eyraud-Dubois, Arnaud Legrand, Martin Quinson, and Frédéric Vivien. A first step towards automatically building network representations. In Proceedings of Euro-Par 2007, volume 4641 of LNCS, pages 160-169, 2007.
- 2 Bobelin, L. and Muntean, T. Algorithms for Network Topology Discovery using End-to-End Measurements. In International Symposium on Parallel and Distributed Computing, 2008 (ISPDC '08), pp.267-274, 1-5 July 2008.
- 3 F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi : a decentralized network coordinate system. Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications, pages 15-26, 2004.
- 4 Ramasubramanian, V., Malkhi, D., Kuhn, F., Balakrishnan, M., Gupta, A., and Akella, A. On the treeness of internet latency and bandwidth. In Proceedings of the Eleventh international Joint Conference on Measurement and Modeling of Computer Systems (Seattle, June 15 - 19, 2009). SIGMETRICS '09, pages 61-72, 2009. <http://doi.acm.org/10.1145/1555349.1555357>
- 5 Shavitt, Y. and Tankel, T. 2008. Hyperbolic embedding of internet graph for distance estimation and overlay construction. IEEE/ACM Trans. Netw. 16, 1 (Feb. 2008), 25-36. <http://dx.doi.org/10.1109/TNET.2007.899021>
- 6 David P. Anderson. Boinc : A system for public-resource computing and storage. In GRID '04 : Proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing, pages 4-10, Washington, DC, USA, 2004. IEEE Computer Society.
- 7 Olivier Beaumont, Lionel Eyraud-Dubois and Shailesh K. Agrawal. Broadcasting on Large Scale Heterogeneous Platforms under the Bounded Multi-Port Model. In 24th IEEE International Parallel & Distributed Processing Symposium (IPDPS'10). To appear.

10 Recherche d'informations et routage par agents mobiles dans des réseaux dynamiques et défailants

Responsable : Nicolas Hanusse

lieu : LaBRI

téléphone : 05.40.00.26.04

e-mail : nicolas.hanusse@labri.fr

équipe : CombAlgo, thème : Algorithmique distribuée, Graphes projet : CEPAGE

Mots-clés : paradigme agent mobile, graphes dynamiques, réseau, routage, tolérance au panne

Description du sujet

Un agent mobile est un programme circulant dans des réseaux et ayant la capacité de s'exécuter quand il arrive à un noeud du réseau. Ici, le terme "réseau" est pris au sens large car il peut inclure aussi bien des réseaux de type Internet, ad hoc, graphe du Web, réseaux pair à pair. Les applications sont multiples : mise-à-jour automatique de données (tables de routage, paramétrages divers, etc.), recherche et collecte d'informations, appels automatiques de numéros de téléphones, apprentissage d'un "plan" d'un environnement/terrain (avec ou sans obstacles), etc.

Nous nous intéressons aux agents mobiles de recherche. Dans un environnement centralisé, un agent va interroger un annuaire ou des tables dont il connaît la localisation. C'est plus ou moins la démarche que fait un utilisateur lorsqu'il interroge un moteur de recherche sur le Web. Cette approche pose plusieurs problèmes :

- certains noeuds doivent contenir énormément d'informations,
- le temps de mise-à-jour des annuaires ou tables lors de changement peut être extrêmement important.

En résumé, cette approche est très gourmande en espace mémoire et en temps de mise-à-jour. D'autre part, le caractère dynamique de certains réseaux rend l'accès à une information existante non encore indexée très difficile.

Dans ce sujet, nous proposons une approche distribuée : les noeuds contiennent des indications partielles pour trouver des informations. Ce type de méthode est particulièrement adapté à des réseaux dynamiques ou pair à pair dans lesquels tous les noeuds possèdent un même rôle, contrairement à l'approche centralisée.

L'objectif de ce mémoire est de concevoir, étudier et expérimenter des algorithmes menant à des informations avec différentes contraintes :

- le réseau peut être défailant. Il est intéressant de construire des algorithmes auto-stabilisants ;
- la quantité de mémoire des noeuds ou des agents peut être bornée ;

– etc.

Notamment, nous nous focaliserons sur des alternatives au protocole BGP utilisé dans les réseaux.

Le stage débutera par une étude bibliographique des principaux travaux liés à ce sujet. Le stagiaire devra ensuite développer des modèles réalistes pour des systèmes d'agents mobiles de recherche. Des algorithmes seront conçus pour le modèle le plus approprié. L'exactitude et l'efficacité des algorithmes seront validés par analyse théorique et/ou par des simulations.

Co-encadrants potentiels : Ralf Klasing, David Ilcinkas

Références bibliographiques

– <http://www.labri.fr/hanusse>

11 Calcul distribué de spanners en grande dimension

Responsable : Nicolas Hanusse

lieu : LaBRI

téléphone : 05.40.00.26.04

e-mail : nicolas.hanusse@labri.fr

équipe : CombAlgo, thème : Algorithmique distribuée, Graphes projet : CEPAGE

Mots-clés : masses de données, résumés de graphes, calcul distribué

Description du sujet

Depuis quelques années il existe des systèmes distribués permettant d'estimer à l'avance la latence (le RTT) entre 2 machines d'un réseau (internet par exemple). Le principe est le suivant : on associe à chaque machine m un point $pt(m)$ d'un espace donné de dimension d possiblement élevée. L'estimation de la latence entre 2 machines m_1 et m_2 est alors la distance entre les points $pt(m_1)$ et $pt(m_2)$ dans l'espace considéré.

L'interconnexion réelle des noeuds dans de tels systèmes est souvent complexes et il est alors courant de considérer un réseau virtuel de communication. Les graphes complets peuvent paraître comme idéaux comme topologie de réseaux de communication mais comportent néanmoins de nombreux inconvénients : synchronisation délicate si les degrés sont élevés, puissance nécessaire s'il s'agit de réseau radio, diffusion coûteuse, etc. Les spanners sont des sous-graphes peu denses, possédant peu d'arêtes, de degré maximum borné, de faible poids et qui approximent bien les distances dans le graphe d'origine. Ils permettent ainsi de construire des réseaux virtuels relativement simples, sans ces inconvénients, mais conservant les propriétés de distance/latence du réseau initial. Les spanners sont aussi très utiles

pour la visualisation de grands graphes pour lesquels on cherche à supprimer la manipulation de nombreuses arêtes. Les spanners peuvent également être utilisés pour répondre à des requêtes de plus proche point, utiles entre autres en fouille de données.

Ce stage porte sur la construction de spanners pour des graphes géométriques : les noeuds possèdent des coordonnées dans le plan (et plus généralement dans un espace de dimension d). La pondération des arêtes correspond à la distance euclidienne. Enfin la longueur du chemin entre toute paire de noeuds dans le spanner doit être une bonne approximation de la distance euclidienne entre ces deux noeuds. Le maximum sur toutes les paires de noeuds de ce facteur d'approximation est appelé l'étirement du spanner.

L'objectif de ce stage est de trouver une construction (a priori distribuée) d'un spanner en dimension d tel que le degré moyen et l'étirement soient tous deux polynomiaux en d . Au moins l'un des deux paramètres est en effet exponentiel en la dimension dans les approches existantes.

Le stage débutera par une étude bibliographique de spanners incluant les triangulations de Delaunay, les graphes de Yao, les theta-graph, etc. Le stagiaire pourra ensuite proposer différentes approches de construction de spanners qu'il validera ou infirmera grâce à une analyse rigoureuse de l'étirement et du degré moyen, ou de façon empirique par le développement d'un petit simulateur, voire des deux façons, en fonction de ses goûts et de ses compétences.

Co-encadrants potentiels : Nicolas Bonichon, David Ilcinkas, Ralf Klasing

Références bibliographiques

- Geometric Spanner Networks, Narasimhan, Giri ; Smid, Michiel, 2007, Cambridge University Press
- http://www.cl.cam.ac.uk/teaching/2003/AdvSysTop/InternetCoordinateSystems_cl03_2004.ppt
- <http://www.labri.fr/hanusse/>

12 Algorithmique des agents mobiles

Responsable : David Ilcinkas

lieu : LaBRI

téléphone : 05.40.00.69.12

e-mail : david.ilcinkas@labri.fr

équipe : CombAlgo, thème : Algorithmique distribuée, Graphes projet : CE-PAGE

Mots-clés : agents mobiles, robots, graphes dynamiques, tolérance aux pannes.

Description du sujet

Un agent mobile est une entité autonome capable de se déplacer dans son environnement. Nous nous intéressons ici à des agents mobiles se déplaçant dans des réseaux, pouvant être virtuels (réseaux informatiques par exemple) ou réels (galeries souterraines, réseaux de tuyauterie, etc.). Dans le premier cas, un agent mobile est un agent logiciel pouvant s'exécuter sur la machine sur laquelle il se situe. Dans le second cas, l'agent mobile est un robot réel/physique.

La dernière décennie a vu un intérêt croissant dans l'utilisation d'agents mobiles dans les environnements distribués. De nombreux systèmes d'agents mobiles logiciels ont été proposés pour les réseaux informatiques et plusieurs ont été totalement implémentés. Ces systèmes permettent la création aisée et le transfert de code mobile dans un réseau mais ne fournissent aucune primitive algorithmique non-triviale et efficace. Comme Nicola Santoro [1] l'a noté et souligné dans l'un des deux exposés invités de la conférence SIROCCO 2008, les systèmes d'agents mobiles manquent de fondements théoriques solides et leur puissance de calcul est peu connue. Le calcul par agents mobiles a cependant reçu un intérêt croissant ces dernières années, en particulier dans la communauté du calcul réparti. Des chercheurs commencent à étudier la robustesse et la puissance des systèmes d'agents mobiles et à concevoir des algorithmes dédiés à ce cadre. Ce sujet de stage s'inscrit dans cette thématique en développement.

L'objectif du stage est d'étudier un ou plusieurs problèmes classiques d'agents mobiles (tels que l'exploration, le rendez-vous, la capture d'un intrus) en portant un intérêt particulier à la tolérance aux pannes et/ou au dynamisme, en fonction des goûts et compétences du candidat.

Co-encadrants potentiels : Nicolas Hanusse, Ralf Klasing

Références bibliographiques

- Mobile Entities Computing : Models and Problems. In the Proceedings of the 13th International Colloquium on Structural Information and Communication Complexity (SIROCCO), volume 5058 of LNCS, page 1, 2008.
- <http://www.labri.fr/~ilcinkas/>

13 Stratégies d-sures de téléchargement pour les systèmes P2P en cas de clients inciviques

Responsable : Colette Johnen

lieu : LaBRI

téléphone : 05.40.00.60.47

e-mail : colette.johnen@labri.fr

équipe : Combinatoire et Algorithmique, thème : algorithme distribué, projet : stratégies d-sures de téléchargement pour les systèmes P2P en cas de

clients inciviques

Mots-clés : système P2P, partage de fichier, gestion du téléchargement

Description du sujet

Initialement, un document (vidéo, musique, etc.) est détenu par une source, des clients téléchargent ce documents à partir de la source. Plus le document est demandé à la source plus l'unique serveur sature, une file d'attente de demande se crée, jusqu'au point de saturation. Le protocole BitTorrent renverse cette tendance en obligeant chaque client informatique ayant téléchargé un segment du document à devenir aussitôt un poste serveur à son tour (c.a.d qu'il accepte le téléchargement du segment à partir de lui). Le principe est alors assez comparable à celui de la réaction en chaîne en physique nucléaire : les téléchargements d'un document s'effectuent de poste à poste sans faire appel à la source initiale du document.

Les techniques utilisées sont : le morcellement du fichier en blocs et le téléchargement depuis différents pairs "peer" des blocs d'un même fichier (on nomme cette technique "multisourcing"). Les blocs peuvent arriver dans un ordre quelconque depuis des sources multiples, le fichier étant réputé téléchargé lorsque la totalité des blocs est parvenue à destination, quel que soit l'ordre d'arrivée de ceux-ci ou leurs provenances.

Il existe plusieurs méthodes pour inciter les postes client à jouer le rôle de serveur. Mais une fois, un poste client satisfait (ayant téléchargé tous les blocs du document), il peut sans mesure de rétention quitter le système Pair à Pair. Un tel poste est qualifié d'incivique.

Une stratégie est d-sure si un client télécharge d copies de chaque bloc à partir de sources distinctes, c'est seulement après avoir téléchargé les d copies du bloc et trouvé, si nécessaire, la bonne version du bloc que le client peut devenir «source/serveur» de ce bloc. Ce type de stratégies tolère que certains postes s'arrêtent au milieu du téléchargement du document, où que des postes aient des blocs corrompus.

Le but de ce stage est de concevoir des stratégies d-sures pour une valeur de $d > 2$. Par exemple, déterminer le taux de demandes de téléchargement maximal que peut gérer une stratégie d-sure ; puis de concevoir cette stratégie.

<http://dept-info.labri.fr/~johnen>

Remarque : Une allocation doctorale sur le même sujet sera proposée.

Références bibliographiques

- "Strategies for Peer-to-Peer Downloading", Ted Herman, Colette Johnen, IPL, 94(5) :203-209, 2005
- "Safe Peer-to-Peer Self-Downloading", Ghosh Dastidar, Ted Herman, Colette Johnen, SSS'06

14 Partitions de l'ensemble des arêtes d'un graphe

Responsable : Mickaël Montassier, André Raspaud

lieu : LaBRI

téléphone : 05.40.00.69.86

e-mail : mickael.montassier@labri.fr, raspaud@labri.fr

équipe : Combinatoire et Algorithmique, thème : Graphes et Applications

Mots-clés : Graphes, partitions

Description du sujet

D'après un résultat de Nash-Williams [NW64], tout graphe planaire $G = (V, E)$ admet une partition des arêtes en trois forêts, dont une est de degré maximum au plus 4 [Gon09], i.e. tout graphe planaire admet une partition de l'ensemble de ses arêtes E en trois ensembles E_1, E_2, E_3 , tels que chacun des sous-graphes G_{E_i} ayant pour ensemble d'arêtes E_i est une forêt ; de plus, il existe au moins un i tel que le degré maximum de G_{E_i} est au plus 4. Si la maille (taille d'un plus petit cycle) du graphe planaire G est au moins 9 (resp. 7, 5), alors il admet une partition des arêtes en une forêt et un sous-graphe de degré maximum au plus 1 [BK+08] (resp. 2 [HH+02], 4 [HH+02]). Gonçalves a montré également que tout graphe planaire de maille au moins 6 (resp. 7) admet une partition des arêtes en deux forêts dont une est de degré maximum au plus 4 (resp. 2) [Gon09].

L'optimalité de ces résultats n'est pas connue :

- Existe-t-il k tel que tout graphe planaire de maille au moins 5 admet une partition des arêtes en deux forêts dont une a un degré maximum au plus 5 ? Si k existe, quelle est la plus petite valeur de k ?
- Quelle est la plus petite valeur de k pour laquelle tout graphe planaire de maille au moins 6 (resp. 7, 8) admet une partition des arêtes en deux forêts dont une a un degré maximum au plus k ?

D'autres partitions ont été étudiées (forêts d'étoiles, forêts linéaires, "caterpillar", forêts de bi-étoiles) dans le cadre des graphes planaires et des graphes planaires de grandes mailles.

Récemment, il a été montré que tout graphe, dont le degré moyen maximum est borné par $\frac{4(k+1)(k+3)}{k^2+6k+6}$, admet une partition des arêtes en une forêt et un sous-graphe de degré maximum au plus k [MPRWZ09]. Il serait intéressant de poursuivre cette étude sur les graphes peu denses (faible degré moyen maximum) en prenant en compte d'autres partitions (forêts d'étoiles, forêts linéaires, "caterpillar", forêts de bi-étoiles, etc.).

Ce type de décomposition a bien évidemment un intérêt structurel et est de plus un outil pour majorer certains invariants de graphe, comme le nombre chromatique du jeu de marquage [Zhu99], ou comme le rayon spectral [DM09].

Dans un premier temps, il faudrait réaliser un état de l'art sur l'ensemble des décompositions existantes et exhiber les liens entre ces différentes décompositions. Dans un second temps, il faudrait essayer de répondre aux questions d'optimalité présentées ci-dessus.

- BK+08 O.V. Borodin, A.V. Kostochka, N.N. Sheikh, and G. Yu. Decomposing a planar graph with girth 9 into a forest and a matching. *European Journal of Combinatorics*, 29(5) :1235–1241, 2008.
- DM09 Z. Dvořák and B. Mohar. Special radius of finite and infinite planar graphs and of graphs of bounded genus. University of Ljubljana, Preprint series, 47 :1096, 2009. <http://www.imfm.si/preprinti/PDF/01096.pdf>
- Gon09 D. Gonçalves. Covering planar graphs with forests, one having bounded maximum degree. *Journal of Combinatorial Theory, Series B*, 99 :314–322, 2009.
- HH+02 W. He, X. Hou, K.-W. Lih, J. Shao, W. Wang, and X. Zhu. Edge-partitions of planar graphs and their game coloring numbers. *Journal of Graph Theory*, 41 :307–317, 2002.
- MPRWZ09 M. Montassier, A. Pêcher, A. Raspaud, D.B. West, and X. Zhu. Decomposition of sparse graphs, with application to game coloring number. Manuscript, 2009.
- NW64 C.St.J.A. Nash-Williams. Decompositions of finite graphs into forests. *J. London Math. Soc.*, 39(12), 1964.
- Zhu99 X. Zhu. The game coloring number of planar graphs. *Journal of Combinatorial Theory, Series B*, 75 :245–258, 1999.

15 Homomorphisme de graphes

Responsable : André Raspaud

lieu : LaBRI

téléphone : 05.40.00.69.33

e-mail : raspaud@labri.fr

équipe : Combinatoire et Algorithmique, thème : Graphes et Applications

Mots-clés : Graphes, homomorphismes

Description du sujet

Un homomorphisme d'un graphe $G = (V, E)$ dans un graphe $G' = (V', E')$ est une application h de V dans V' telle que si $xy \in E$ alors $h(x)h(y) \in E'$. Les problèmes de coloration des sommets d'un graphe peuvent être considérés du point de vue des homomorphismes. Il a été prouvé que les sommets d'un graphe sans triangle et ne contenant pas K_5 comme mineur sont 3-coloriables [REZA+09]. Nous

avons prouvé que si le degré moyen maximum d'un graphe G est strictement plus petit que $\frac{5}{2}$ alors il existe un homomorphisme de G dans le graphe de Petersen. Le sujet de ce mémoire est, dans un premier temps, de faire un état de l'art concernant les homomorphismes de graphes sous certaines contraintes (degrés moyen maximum, grande maille, etc.).

Une très belle conjecture de Nešetřil stipule qu'il existe un entier k tel que si un graphe cubique G a une maille plus grande que k alors il existe un homomorphisme de G dans C_5 (le cycle ayant 5 sommets).

On ne demande pas de résoudre cette conjecture, mais on demande de voir, dans un deuxième temps, si avec certaines conditions supplémentaires il existe un tel homomorphisme. Par exemple il a été prouvé que si le degré moyen maximum d'un graphe G est strictement plus petit que $\frac{12}{5}$ alors un tel homomorphisme existe [BK+08].

BK+08 O.V. Borodin, S. G. Hartke, A.O. Ivanova, A.V. Kostochka, D.B. West. Circular $(5, 2)$ -coloring of sparse graphs *Siberian Electronic Mathematical Reports*, (5) :417–426, 2008.

REZA+09 R. Naserasr, Y. Nigussie, R. Škrekovski. Homomorphisms of triangle-free graphs without a K_5 -minor. *Discrete Mathematics*, 309 (2009), 5789–5798.

16 Coloration orientée des graphes de degré au plus trois

Responsable : Éric Sopena

lieu : LaBRI

téléphone : 05.40.00.38.71

e-mail : eric.sopena@labri.fr

équipe : Combinatoire et Algorithmique, thème : Graphes et Applications

Mots-clés : coloration orientée, homomorphismes, graphe orienté, graphe de degré au plus trois.

Description du sujet

Le sujet est décrit à l'adresse suivante.

http://www.labri.fr/perso/sopena/pmwiki/uploads/Proposals/2010_Sopena_M2R_Sujet_2.pdf

17 Coloration d'incidences

Responsable : Éric Sopena

lieu : LaBRI

téléphone : 05.40.00.38.71

e-mail : eric.sopena@labri.fr

équipe : Combinatoire et Algorithmique, thème : Graphes et Applications

Mots-clés : graphe, coloration, coloration d'incidence, classe de graphes

Description du sujet

Le sujet est décrit à l'adresse suivante.

http://www.labri.fr/perso/sopena/pmwiki/uploads/Proposals/2010_Sopena_M2R_Sujet_1.pdf

18 Modèles de communication et diffusion d'information

Responsable : Akka Zemmari

lieu : LaBRI

téléphone : 05.40.00.60.93

e-mail : akka.zemmari@labri.fr

équipe : Combinatoire et algorithmes, thème : Algorithmique distribuée

Mots-clés : Algorithmes distribués, Graphes

Description du sujet

Un système distribué est une collection d'unités physiques ou logiques autonomes – unités ayant leurs propres moyens de contrôle et de calcul – communiquant via des liens de communication. Ces unités peuvent être des ordinateurs ou des processeurs (unités physiques) ou simplement des processus (unités logiques). Il est d'usage de modéliser un système distribué par un graphe $G = (V, E)$ dont les sommets représentent les entités du système et les arêtes les liens de communication.

Dans la littérature, plusieurs modes de communication sont possibles : une entité peut envoyer un message à un ou tous ses voisins, une entité peut recevoir de tous ou partie de ses voisins, deux voisins ne peuvent communiquer que s'ils sont synchronisés, etc.

Un algorithme distribué est un algorithme capable de faire collaborer tous, ou partie, des entités du système afin de réaliser une tâche précise.

Le problème de la diffusion dans les systèmes distribués est l'un des paradigmes fondamentaux de l'algorithmique distribuée. Un processus détient une information et désire la diffuser à tous les autres processus du système. Résoudre ce problème consiste à concevoir un algorithme (déterministe ou probabiliste) pour réaliser une telle diffusion en utilisant le mode de communication utilisé dans le système.

L'objectif de ce travail est :

- faire une (brève) étude bibliographique du sujet, et comparer les différentes solutions proposées en fonction des modes de communication adoptés,
- analyser et proposer une solution probabiliste pour réaliser les modes de communication listés ci-dessus et analyser la diffusion dans chacun de ces modes de communication,
- programmer et tester les solutions proposées en utilisant la plate-forme VISIDIA.

<http://www.labri.fr/~zemmari> <http://visidia.labri.fr>

Références bibliographiques

- Yves Métivier, Nasser Saheb and Akka Zemmari. Analysis of a Randomized Rendez Vous Algorithm. *Information and Computation*, 184 :109-128, 2003.
- Philippe Duchon, Nicolas Hanusse, Nasser Saheb and Akka Zemmari. Broadcast in the Rendezvous Model. *Information and Computation*. 204 :697-712, 2006.

Combinatoire et Algorithmique - MABioVis

19 Ensembles fréquents maximaux

Responsables : Romain Bourqui et Nicolas Hanusse

lieu : LaBRI

téléphone : 05.40.00.26.04

e-mail : romain.bourqui@labri.fr

équipe : MaBioVis/CombAlgo, thème : Visualisation, Graphes projet : GRAVITE/CEPAGE

Mots-clés : fouille de données, visualisation, algorithmes de graphes

Description du sujet

Un des algorithmes de références en fouille de données porte sur le calcul des ensembles fréquents maximaux. Une *transaction* est constituée d'une liste d'items et une *table de transactions* comporte des milliers de transactions. L'objectif est de découvrir quelles sont les combinaisons d'items les plus fréquentes qui apparaissent dans une transaction.

Par exemple, une transaction peut représenter un ticket de caisse d'un magasin et la table de transactions peut constituer l'ensemble des tickets de caisse de l'année. Ainsi, il a été montré en utilisant un calcul de fréquents maximaux que de nom-

breux pères de famille achetaient simultanément bière et couche lors des courses hebdomadaires. De manière plus générale, le calcul de fréquents maximaux peut aider à bâtir des hypothèses sur les caractéristiques de maladies, la découverte des chemins les plus employés dans les réseaux, etc.

Un des obstacles de l'usage de tels algorithmes porte sur l'interaction avec les données : les résultats de calculs de fréquents maximaux sont souvent représentés par de grandes listes et seules les premières entrées sont regardées. D'autre part, le moindre changement de paramètres implique un recalcul des fréquents maximaux et il devient impossible d'effectuer des comparaisons avec les résultats précédents.

L'objectif de ce travail est de proposer des mécanismes d'interaction et visualisation des fréquents maximaux. Par exemple, la représentation en graphe ou coordonnées parallèles pourront être envisagées. Des expérimentations sous Tulip permettront de faire des comparaisons sur des jeux de données réels.

Le site <http://fimi.cs.helsinki.fi/> regorge d'expérimentations et articles sur le calcul des fréquents maximaux.

20 Requêtes skyline et Spanners géométriques

Responsable(s) : Nicolas Hanusse et Sofian Maabout

lieu : LaBRI

téléphone : 05.40.00.26.04 et/ou 05.40.00.35.27

e-mail : {hanusse, maabout}@labri.fr

équipe(s) : CombAlgo & MaBioVis

Mots-clés : Spanner, Skyline, Top-K, Optimisation de requêtes

Description du sujet

Le but de ce sujet est l'étude et la proposition d'algorithmes de requêtes multidimensionnelles de type "skyline", ou d'une manière équivalente, Top-k en base de données. Informellement, un graphe est construit à partir de relations entre les données et toute requête correspond à une navigation dans ce graphe. Malheureusement, les propositions de littérature manipulent de très gros graphes, difficilement maintenable en mémoire et pour lesquels les temps de requêtes ne sont pas garantis.

Etant donné un ensemble de points \mathcal{P} dans un espace multidimensionnel, la requête skyline, notée $\mathcal{S}(\mathcal{P})$ retourne le sous-ensemble de \mathcal{P} qui est constitué de l'ensemble des points qui ne sont *dominés* par aucun autre point : On dit que p_1 domine p_2 si (1) pour chaque dimension, la coordonnée de p_1 est inférieure ou égale à celle de p_2 et (2) pour au moins une dimension, la coordonnée de p_1 est strictement inférieure à celle de p_2 . Par exemple, on cherche les restaurants les moins loins et les moins chers par rapport à un point donné. L'ensemble des restaurants

représente l'ensemble \mathcal{P} dont les coordonnées sont (*Distance, Prix*). La requête Top-k retourne les k meilleurs points non dominés, c'est dans ce sens que Top-k est un cas particulier de Skyline.

L'originalité de l'approche proposée dans cette étude est l'usage de spanners. Les spanners représentent des résumés de graphe permettant de garantir une certaine qualité lors du calcul de distance. Le travail proposé consiste à étudier l'apport de ces derniers pour le calcul des requêtes skyline.

Le mémoire commencera par une étude de bibliographique puis par la conception d'un algorithme basé sur les spanners. Selon la description de l'algorithme obtenu, il pourra être envisagé de réaliser une expérimentation sur des jeux de données réels.

Références bibliographiques

- PREFER : A System for the Efficient Execution of Multi-parametric Ranked Queries International Conference on Management of Data (SIGMOD), 2001, Vagelis Hristidis, Nick Koudas, Yannis Papakonstantinou
- Algorithms and Analyses for Maximal Vector Computation. VLDB Journal. Vol 16, Number 1, January 2007, pages 5-28. Parke Godfrey, Ryan Shipley, Jarek Gryz,
- Dominant Graph : An Efficient Indexing Structure to Answer Top-K Queries, in Proceedings of the 24th International Conference on Data Engineering (ICDE), 2008. Lei Zou, Lei Chen
- Geometric Spanner Networks, Narasimhan, Giri ; Smid, Michiel, 2007, Cambridge University Press

MaBioVis

21 Dessin de graphes dynamiques

Responsables : Guy Melançon et Bruno Pinaud

lieu : LaBRI

téléphone : 05.40.00.37.97

e-mail : bruno.pinaud@labri.fr

équipe : MaBioVis, thème : Visualisation de grandes masses de données,

projet : Porgy

Mots-clés : visualisation de graphes dynamiques, réécriture de graphes

Description du sujet

Ce sujet est une version courte et sans illustration de ceci. Ce stage concerne la visualisation de graphes dynamiques apparaissant dans le cadre de systèmes de réécriture de graphes. La réécriture de graphes apparait aujourd'hui dans des domaines d'applications divers. On peut citer le cas des réseaux d'interactions moléculaires, où les sommets des graphes représentent les molécules qui s'agencent les unes avec les autres au gré des réactions biochimiques décrites par des règles de réécriture précises. Ces règles permettent alors de tenir compte de la réalité biochimique et de contraindre les façons dont les molécules peuvent s'agencer.

Nous avons déjà une première version de la plate-forme de réécriture qui permet à un utilisateur d'appliquer des règles de réécriture sur un graphe, appelé modèle, de conserver un historique des différentes opérations effectuées en construisant notamment un arbre de dérivations. Ses sommets sont les différents états du modèle et une arête relie deux sommets a et b si l'application d'une règle de réécriture permet de transformer a en b . Comme c'est l'utilisateur qui pilote le processus de réécriture, on ne peut donc pas connaître a priori comment vont évoluer le modèle et l'arbre de dérivation. Néanmoins, ces informations sont importantes pour guider les algorithmes de visualisations dynamiques qui doivent entre autres maintenir une certaine stabilité des dessins entre deux tracés consécutifs. De plus, si l'arbre de dérivations devient trop important, certaines parties devront être repliées/compressées de manière (semi-)automatique afin de conserver un tracé lisible et compréhensible.

L'objectif de ce stage est de participer au développement de ces algorithmes de visualisation dynamique (pour le moment, nous utilisons simplement des algorithmes classiques de tracé statique). On s'intéressera plus particulièrement au dessin dynamique du modèle dans un premier temps.

Le sujet détaillé est disponible : [ici](#)

Références bibliographiques

- Frishman, Y. & Tal, A. Online Dynamic Graph Drawing IEEE Trans. on Visualization and Computer Graphics, IEEE Educational Activities Department, 2008, 14, 727-740.
- Branke, J. Kaufmann, M. & Wagner, D. (ed.) Dynamic Graph Drawing Drawing Graphs : methods and models, Springer, 2001, 228-246

22 Dessin de graphes contraints

Responsables : Guy Melançon et Bruno Pinaud

lieu : LaBRI

téléphone : 05.40.00.37.97

e-mail : bruno.pinaud@labri.fr

équipe : MaBioVis, thème : Visualisation de grandes masses de données,
projet : Porgy

Mots-clés : visualisation de graphes dynamiques, réécriture de graphes

Description du sujet

Ce sujet est une version courte et sans illustration de ceci. Le stage proposé concerne le dessin de graphes statiques extrêmement contraint. Généralement, les algorithmes de visualisation travaille à améliorer un ou plus rarement deux critères sans se préoccuper du reste. Nous voulons mettre au point un algorithme qui puisse prendre en compte de nombreux critères concurrents.

Ce travail se place dans le cadre du développement d'un système de réécriture de graphe. La réécriture de graphes apparait aujourd'hui dans des domaines d'applications divers. On peut citer le cas des réseaux d'interactions moléculaires, où les sommets des graphes représentent les molécules qui s'agencent les unes avec les autres au gré de réactions biochimiques décrites par des règles de réécriture précises. Ces règles permettent alors de tenir compte de la réalité biochimique et de contraindre les façons dont les molécules peuvent s'agencer. Les règles de réécriture sont décrites à l'aide de graphes et doivent être correctement dessinées et visualisées pour aider l'utilisateur à raisonner sur le système, ou encore pour lui permettre de sélectionner une règle à appliquer. Elles impliquent deux graphes, un membre gauche (qu'il faut trouver dans le graphe à réécrire), et un membre droit (qui doit remplacer une instance du membre gauche dans le graphe à réécrire), mis en correspondance par une partie médiane. Cette distinction gauche/droite permet de représenter la règle comme un seule et unique graphe. La mise en correspondance des sommets et des arêtes qui doivent être préservés au moment de la réécriture ajoute à cette paire de graphes des arêtes additionnelles qui sont connectées par l'intermédiaire d'une structure médiane appelée "bridge".

L'objectif de ce stage est de mettre au point un algorithme de visualisation efficace et robuste des graphes formés par les règles. La difficulté est que les contraintes associées à ces dessins sont nombreuses et combinent de nombreux aspects de la visualisation :

- Les différentes parties de la règle doivent être à leur place respective (i.e. le membre gauche doit être sur la gauche du dessin).
- Bien souvent, il y a une forte symétrie entre les membres gauche et droit (le membre gauche ressemble beaucoup au membre droit). Cette symétrie doit être préservée et exploitée.

Le dessin de chaque membre doit bien sûr être lisible et compréhensible. Néanmoins, le graphe qui représente la règle entière doit aussi être lisible. Aux critères classiques qui permettent d'améliorer la lisibilité du graphe tel que la minimisation du nombre de croisements d'arêtes s'ajoute le routage des arêtes pour évi-

ter qu'elles ne chevauchent des sommets. Néanmoins, ces critères sont bien souvent contradictoires. L'amélioration d'un premier critère peut fortement dégrader la qualité d'un second produisant au final un dessin de mauvaise qualité. L'algorithme à développer devra trouver le meilleur compromis possible entre tous ces critères pour produire un tracé de la meilleure qualité possible. Un piste possible est de s'inspirer des travaux de T. Dwyer qui utilise la programmation par contraintes pour produire des algorithmes de tracé.

Le sujet détaillé est disponible : [ici](#)

Références bibliographiques

- T. Dwyer, Y. Koren et K. Marriot. IPSep-CoLa : An Incremental Procedure for Separation Constraint Layout of Graphs. *IEEE Trans. on Visualization and Computer Graphics* 12(5) : 821–828, 2006
- T. Dwyer, K. Marriott et M. Wybrow. Integrating Edge Routing into Force-Directed Layout. In *Proc. 14th Intl. Symp. Graph Drawing (GD '06)*, volume 4372 of *Lecture Notes in Computer Science*, pages 8–19. Springer, 2007.
- T. Dwyer, K. Marriott, F. Schreiber, P. J. Stuckey, M. Woodward et M. Wybrow. Exploration of networks using overview+detail with constraint-based cooperative layout. *IEEE Trans. on Visualization and Computer Graphics*, 14(6) :1293–1300, 2008
- Freivalds, K. Goos, G. ; Hartmanis, J. & van Leeuwen, J. (ed.) *Curved Edge Routing 13th Int. Symp. on Fundamentals of Computation Theory*, Springer, 2001, 2138, 126–137
- Gansner, E. R. ; Koutsofios, E. ; North, S. C. & Vo, K.-P. A Technique for Drawing Directed Graphs *IEEE Trans. on Software Engineering*, 1993, 19, 214–230

Image et Son

23 Analyse des réseaux métaboliques

Responsable : Marie Beurton-Aimar

lieu : LaBRI

téléphone : 05.40.00.35.25

e-mail : beurton@labri.fr

Mots-clés : graphes, chemins minimaux, résolution de contraintes.

Description du sujet

Les réseaux métaboliques sont des graphes ayant deux types de noeuds, l'un

pour les molécules impliquées, l'autre pour la représentation des processus (réactions biochimiques). Ces graphes bien que de taille moyenne (en général moins d'une centaine de noeuds) sont considérés comme des graphes *petit-monde* en ceci qu'un petit nombre de noeuds forment des hubs entre des parties (sous graphes) très connectées.

Pour analyser les interactions entre les différents objets biologiques appartenant au réseau, il est intéressant d'identifier les chemins empruntés par les molécules dans ce graphe. Il existe des méthodes comme les modes élémentaires [1] qui permettent de détecter ces chemins qui respectent un ensemble de contraintes biologiques [2]. Toutefois, cette méthode basée sur des méthodes d'analyse linéaire, peut rencontrer des problèmes calculatoires, dûs au nombre de chemins potentiellement valides (en général plusieurs milliers pour un graphe d'une centaine de noeuds) [3].

Le stage proposé consiste premièrement en une analyse des limites de la méthode des modes élémentaires pour éventuellement proposer une amélioration de l'algorithme de calcul (cf [4]). Deuxièmement, une étude de faisabilité autour des systèmes de réécriture est demandée, les réseaux métaboliques étant des réseaux de réactions biochimiques, ils peuvent être vus comme un ensemble de règles de transformations (réécritures) de termes (molécules impliquées).

Pour ce sujet nous disposons de plusieurs jeux de données qui serviront d'applications concrètes pour tester les différentes solutions proposées. Ce travail ne requière aucune connaissance en biologie car il est mené en collaboration avec une équipe de l'INRA (modélisation du métabolisme du fruit) qui assure l'interprétation au niveau biologique des travaux théoriques.

Références bibliographiques

- 1 http://en.wikipedia.org/wiki/Metabolic_network_modelling
- 2 <http://www.csb.ethz.ch/tools/efmtool>
- 3 "*Elementary mode analysis : a useful metabolic pathway analysis tool for characterizing cellular metabolism*" Cong T. Trinh Aaron Wlaschin and Friedrich Sreenc, Applied Microbiology and Biotechnology, Volume 81, Number 5 / janvier 2009
- 4 Large-scale computation of elementary flux modes with bit pattern trees Marco Terzer and Jörg Stelling, Bioinformatics, Volume 24, Number 19, Pp. 2229-2235

24 Segmentation 3D par Watershed Inter-Voxel

Responsables : Fabien Baldacci et Luc Brun

Mots-clés : courbure, analyse multi-echelle, matching, interpolation, animation 2D/3D

Description du sujet

Les outils actuels d'animation en informatique graphique, qu'ils soient spécialisés pour des personnages ou objets en 3D ou en 2D, demandent beaucoup de travail et sont réservés à des artistes confirmés. En animation 3D, la tâche consiste par exemple à manipuler les articulations d'un personnage de manière méticuleuse, ou à contrôler les mouvements des muscles du visage pour créer des expressions convaincantes. En animation 2D, le travail s'opère à l'inverse : les artistes doivent imaginer la succession de dessins qui va pouvoir donner lieu à un mouvement expressif, en considérant le poids des différentes parties du corps. Ces deux tâches sont très différentes et sont donc réservées à des utilisateurs experts dans chacun des deux domaines.

Récemment, de nouvelles techniques d'animation ont été proposées pour faciliter la manipulation de modèles 2D [Igarashi05] ou 3D [Gal09]. Ces outils se fondent sur la manipulation de courbes 2D ou 3D et permettent des éditions intuitives. Leur utilisation pour la création d'animations comporte un défaut majeur cependant : les courbes restent difficiles à déformer de manière intuitive, ce qui limite grandement les possibilités créatrices.

Nous proposons ici de mettre en oeuvre une analyse approfondie d'une courbe de contrôle, qui permette d'identifier des sections caractéristiques comme des coins ou arrondis, et de les manipuler via des interfaces adaptées. Pour cela, nous nous intéresserons dans un premier temps à extraire des descriptions de la courbe en chaque point, et à plusieurs échelles ; puis dans un deuxième temps nous combinerons ces informations afin d'identifier un petit nombre de "widgets" de contrôle.

Une fois ces widgets extraits et en fonction du temps disponible, nous pourrions expérimenter quelques applications : la manipulation intuitive de courbes, en manipulant directement les widgets via une souris, un stylet ou un pad multi-touch ; l'interpolation de deux courbes, en interpolant via les widgets ; voire même la stylisation de courbes, en exploitant les informations haut niveau portées par les widgets.

Références bibliographiques

- Igarashi05 Takeo Igarashi, Tomer Moscovich, John F. Hughes, As-Rigid-As-Possible Shape Manipulation ACM Transactions on Computer Graphics (Proceedings of SIGGRAPH 2005)
- Gal09 Ran Gal, Olga Sorkine, Niloy Mitra, Daniel Cohen-Or, iWIRES : An Analyze-and-Edit Approach to Shape Manipulation, ACM Transaction on Graphics (Proceedings of SIGGRAPH 2009)

26 Localisation par vidéo portée

Responsable : Jenny Benois-Pineau, Svebor Karaman

lieu : LaBRI

téléphone : 05.40.00.84.24, 05.40.00.38.80

e-mail : jenny.benois@labri.fr, svebor.karaman@labri.fr

équipe : Image et son, thème : Analyse et Indexation Vidéo, projet : IMMED

Mots-clés : Vidéo portée, Localisation, Détection d'objets

Description du sujet

Le projet est proposé dans le contexte d'une recherche en analyse et indexation vidéo pour aider les traitements des maladies liées à l'âge. Les médecins ont à disposition la vidéo enregistrée par des caméras portées par des patients et les informaticiens doivent analyser et résumer automatiquement les grandes quantités de vidéos acquises. Un ensemble d'outils pour répondre à ces tâches est en cours de développement au sein du thème "Analyse et Indexation Vidéo" de l'équipe "Image et Son" du LaBRI. Dans le projet, il s'agit de développer un outil d'identification des lieux caractéristiques d'une activité quotidienne d'un patient. Le but sera de détecter dans des images caractérisant des activités des patients (comme laver la vaisselle, préparer le repas, lire, etc.) des éléments caractéristiques comme un évier, une table, une gazinière, un livre, etc. Nous faisons appel ici aux descripteurs locaux tels que les points SURF disponible dans la librairie OpenCV ou semi-locaux basés sur des blocs couleurs extraits de l'image. Il s'agit de mettre au point, en utilisant un algorithme de clustering déjà disponible au LaBRI, un outil de construction de dictionnaires visuels. Ainsi chaque objet (évier, gazinière, table, etc.) sera représenté par son propre dictionnaire. Un travail algorithmique et expérimental consisterait également à réduire les dictionnaires visuels afin de séparer les objets au mieux. Dans la phase de reconnaissance, les mêmes éléments seront extraits des images de la vidéo et on cherchera à déterminer si une activité peut être définie par ces éléments par des méthodes de recherche de plus proches voisins.

Site du projet IMMED : <http://immed.labri.fr>

Références bibliographiques

- Librairie OpenCV
<http://opencv.willowgarage.com/wiki/>
- H. Bay, A. Ess, T. Tuytelaars and L. V. Gool : "SURF : Speeded Up Robust Features", Computer Vision and Image Understanding, 2008.
<http://www.vision.ee.ethz.ch/~surf/eccv06.pdf>
- J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman : "Lost in Quantization : Improving Particular Object Retrieval in Large Scale Image Databases", Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition, 2008.

<http://www.robots.ox.ac.uk/~vgg/publications/papers/philbin08.pdf>

- R. Megret, D. Szolgay, J. Benois-Pineau, P. Joly, J. Pinquier, J.-F. Dartigues, C. Helmer : "Wearable video monitoring of people with age dementia : Video indexing at the service of helthcare", International Workshop on Content-Based Multimedia Indexing, 2008.

http://hal.archives-ouvertes.fr/docs/00/34/91/62/PDF/CBMI_monitoringIntegrated2_accept.pdf

- E. Dumont and B. Mérialdo : "Redundancy removing and event clustering for video summarization", 9th International Workshop on Image Analysis for Multimedia Interative Services, 2008.

<http://www.eurecom.fr/util/publidownload.fr.htm?id=2461>

27 Extraction de descripteurs d'objets vidéo HD dans des pyramides de transformées hiérarchiques

Responsables : Jenny Benois-Pineau, Anne Vialard

lieu : LaBRI

téléphone : 05.40.00.84.24, 05.40.00.69.35

e-mail : jenny.benois@labri.fr, anne.vialard@labri.fr

équipe : Image et son

thèmes : Analyse et Indexation Vidéo / Structuration et Analyse d'image

Mots-clés : vidéo numérique compressée, indexation/recherche, géométrie discrète

Description du sujet

Les techniques d'indexation et de recherche appliquées à des contenus vidéo numériques compressés sont devenus particulièrement nécessaires avec la forte croissance des volumes de ce type d'informations. L'arrivée de la vidéo HD dans les archives et dans la télédiffusion professionnelle a mis en avant le nouveau standard JPEG2000. Ce dernier comporte différents niveaux de résolution des images vidéo représentées par des pyramides d'ondelettes.

Dans un objectif de recherche de vidéos par objet dans des bases de données vidéo à grande échelle, une description/caractérisation est nécessaire. Au sein de l'équipe Image et Son (thème AIV) un nouveau descripteur des objets vidéo hiérarchiques a été proposé. L'extraction de ce dernier demande une segmentation préalable en objets et un calcul de caractéristiques statistiques sur le masque de l'objet à différentes résolutions de la pyramide [1]. Ce descripteur s'avère plus performant que les descripteurs SIFT qui à ce jour constituent l'état de l'art en indexation et recherche des images/vidéo.

Dans le calcul des divers descripteurs de type SIFT, la prise en compte de l'éloignement des éléments de l'image relativement aux centres des régions d'intérêt consiste en une pondération par des fonctions de lissage continues, gaussiennes par exemple. Dans le cadre de ce projet de recherche nous souhaitons étudier les possibilités de lissage discret, développé dans [2]. Ce dernier nous semble prometteur aussi bien sur le plan de prise en compte des phénomènes d'aliasing caractéristiques de la décomposition pyramidale que pour une évaluation plus fine de la contribution des coefficients des ondelettes.

Environnement de travail : C, C++, Linux.

Références bibliographiques

- J. Benois-Pineau and S. Anthoine and C. Morand and E. Debreuve and J.-Ph. Domenger and W. Bel Haj Ali and P. Piro, Scalable Indexing of HD Video, chapitre du livre "High-quality visual experience : creation, processing and interactivity of high-resolution and high-dimensional video signals", Ed. M. Mrak, M. Grgic, M. Kunt. Springer, à paraître
- J.-P. Braquelair and A. Vialard, Euclidean Paths : A New Representation of Boundary of Discrete Regions, Graphical Models and Image Processing, 61, 16-43, 1999

28 Métapiano

Responsable : Myriam Desainte-Catherine, Antoine Allombert et Mauricio Toro

lieu : LaBRI

téléphone : 05.40.00.69.26

e-mail : myriam@labri.fr

équipe : Image et son, thème : Modélisation du son et de la musique

Mots-clés : contraintes, interprétation et partitions

Description du sujet

Le métapiano (Jean Haury/Les Catamurons) est un instrument de musique concentré en neuf touches de piano seulement. Il peut être joué par quelques doigts, voire avec un seul. Le métapiano va au delà d'un instrument traditionnel puisqu'il intègre les notes qui vont être interprétées par le musicien. En pratique, la partition est analysée pour établir une structure interactive réunissant relations mélodiques, harmoniques et contrapunctiques entre les notes. Les hauteurs des notes seules sont codées et mémorisées selon les règles de la pianotechnie en une partition/structure ne prenant en compte que ces relations de hauteur des notes, éliminant tout aspect temporel et dynamique. La partition codée peut alors être interprétée et mise en sons en jouant sur un nombre restreint de touches du métapiano. Le musicien

qui connaît la pièce «d'oreille» lui applique rythmes, tempo, articulations, accents, phrasés dynamique et agogique pour l'interpréter dans son propre style.

Les partitions interactives représentent une généralisation du concept du métapiano. Leur formalisme est basé sur la notion d'événements statiques et interactifs qui peuvent être liés par des relations temporelles au moment de l'écriture. Ces relations seront vérifiées au moment de l'exécution de la pièce. Le modèle d'exécution d'une partition interactive est un réseau de Petri.

Contexte : Le métapiano permet de faire jouer un seul musicien à la fois. Le logiciel est écrit en MAX et nécessite une analyse manuelle préalable. A l'inverse, les partitions interactives fournissent un cadre dans lequel il est possible de spécifier un nombre quelconque de musiciens. La bibliothèque lib-iscore permet d'exécuter une partition interactive qui a été préalablement construite au moyen d'une interface. Plusieurs interfaces sont disponibles (virage, gsharp, acousmoscribe). L'objectif de ce projet est la création de modes de jeux pour les partitions interactives basés sur les travaux de Jean Haury avec le metapiano pour parvenir à la définition d'un métaquatuor.

Travail à réaliser : Il faut réallier une étude de l'importation de partitions du métapiano pour un joueur dans le logiciel I-score. La partition métapiano est composée de notes (hauteurs) et de règles temporelles concernant la succession, la coïncidence et le retard des événements de déclenchements et de relâchements de ces notes. La partition i-score induite sera donc composée de boîtes correspondant aux notes, de relations temporelles issues des règles sur les événements, et de points d'interaction correspondant au jeu inscrit dans la partition métapiano.

Il s'agira ensuite de concevoir des modes de jeu. De la version à un joueur, il faudra déduire les modes de jeu à plusieurs joueurs, ce qui se concrétisera dans I-score par une affectation différente des points d'interaction et des relations temporelles. Les différentes voix pourront être regroupées chacune dans des boîtes hiérarchiques. Il sera nécessaire de mener une étude afin d'aboutir à une stratégie de transformations selon les cas de figure. En effet, afin de préserver la justesse de l'interprétation, la pose des événements interactifs devra se faire après une analyse précise de la configuration des événements de déclenchement et de relâchement de chacune des voix.

Pré-requis : informatique de base (automates, graphes, contraintes), connaissance du langage C ou Lisp et d'un peu de solfège.

lien vidéo <http://hal.archives-ouvertes.fr/hal-00307925/fr/> <http://hal.archives-ouvertes.fr/hal-00353601/>

29 Analyse et comparaison des propriétés timbrales d'un morceau de musique, application à la musique électroacoustique

Responsables : Myriam Desainte-Catherine, Pierre Hanna et Pierrick Legrand (IMB)

lieu : LaBRI

téléphone : 05.40.00.35.07

e-mail : myriam@labri.fr

équipe : Image et Son, thème : Modélisation du Son et de la Musique

Mots-clés : Descripteurs du timbre, similarité musicale, analyse du son

Description du sujet

Le projet ANR SIMBALS du LaBRI s'intéresse à l'évaluation automatique de la similarité entre morceaux de musique. Ce problème ouvert est assez complexe et repose sur de nombreux facteurs tels que le rythme, l'harmonie, la mélodie, etc. Les systèmes existants considérant la similarité timbrale prennent en compte des descripteurs bas-niveau (par exemple MFCC). Nous souhaitons proposer des descripteurs de plus haut-niveau liés à des propriétés musicales, en nous basant sur les travaux reconnus de certains compositeurs de musique électroacoustique.

Les applications sont très nombreuses et concernent aussi bien la transcription automatique de musique électroacoustique, que l'estimation automatique de similarité timbrale ou encore l'analyse de genre ou de style musical.

Projet Simbals

Références bibliographiques

- Premiers résultats pour la similarité métrique
- Premiers résultats pour la similarité mélodique
- Article de Jean-Louis Di-Santo
- Travaux sur les descripteurs audio dans le cadre de Cuidado

30 Système hybride de capture de geste sportif

Responsable : Jean-Christophe Gonzato, Jean-Sébastien Franco

lieu : LaBRI

téléphone : 05.53.48.06.21

e-mail : gonzato@labri.fr

équipe : Image et Son, thème : MVI3D, projet : IPARLA

Mots-clés : Capture Vidéo, système hybride, sport

Description du sujet

Dans de plus en plus de sports, le recours à la vidéo est devenu indispensable dans le processus d'amélioration des résultats. Cela passe de l'étude de certains paramètres du compétiteur (posture, vitesse, positionnement, etc.) à l'analyse d'une stratégie de jeu de l'équipe adverse (tactiques souvent employées, positionnements des joueurs, fatigue, etc.). Dans la plupart des cas, l'oeil de l'expert ou du coach sur un simple enregistrement vidéo suffit pour améliorer la technique du ou des compétiteurs.

Nous proposons dans le cadre de ce mémoire d'établir un état de l'art des systèmes de capture utilisant soit la vidéo soit des capteurs de mouvement positionnés sur le sportif lui-même soit un composé des deux systèmes. Plusieurs aspects devront être présentés :

- hétérogénéités des capteurs
- Synchronisation des capteurs
- Gestions des données provenant des vidéos et des capteurs
- Reconstruction des positions 3D
- Vitesse des objets capturés
- Modes de représentation des informations capturées
- etc.

De plus, dans ce travail, il sera demandé d'établir un premier prototype de capture de mouvement de sportif utilisant un système hybride dans des conditions idéales (salle dédiées, mouvement simple, éclairage contrôlé, etc.).

Références bibliographiques

- MAO Jinzi, Tracking a Tennis Ball Using Image Processing Techniques, Thèse, University of Saskatchewan
- Xiaofeng Tong et al., A Three-Level Scheme for Real-Time Ball Tracking, MCAM'07
- Andrew W. B. Smith and Brian C. Lovell, Visual Tracking for Sports Applications, APRS'05
- Nikos Katsarakis et al., Person tracking for ambient camera selection in complex sports environments, DIMEA'08

31 Interfaces utilisateur pour de la création d'objets 3D

Responsable : Martin Hachet et Christophe Schlick

lieu : LaBRI

téléphone : 05.40.00.69.20

e-mail : [hachet|schlick]@labri.fr

équipe : Image et Son, thème : MVI3D, projet : Iparla

Mots-clés : Interfaces utilisateur, interaction 3D, widgets, modélisation géométrique

Description du sujet

Depuis l'introduction des "Widgets 3D" génériques pour la manipulation et l'édition d'objets graphiques il y a une vingtaine d'années, les interfaces utilisateurs dédiées aux outils de modélisation 3D ont très peu évolué. Ces interfaces sont destinées à une utilisation experte et, par conséquent, ne permettent pas une prise en main aisée par le grand public.

L'objectif de ce projet est de permettre à des utilisateurs novices de créer et d'éditer facilement des formes 3D. Pour cela nous explorerons de nouvelles approches permettant d'obtenir rapidement des formes grossières tout en offrant des possibilités de raffinement au moyen de widgets spécialisés.

Au cours de ce stage de recherche, nous étudierons les approches existantes dans le domaine des interfaces et de la modélisation géométrique et nous mettrons au point de nouvelles techniques que nous testerons au travers d'études utilisateur.

<http://iparla.labri.fr> <http://www.labri.fr/~hachet>

Références bibliographiques

- Ryan Schmidt, Karan Singh, Ravin Balakrishnan. "Sketching and Composing Widgets for 3D Manipulation" (2008) Computer Graphics Forum, 27(2), pp. 301-310. (Proceedings of Eurographics 2008).
- Andrew Nealen, Takeo Igarashi, Olga Sorkine and Marc Alexa, "FiberMesh : Designing Freeform Surfaces with 3D Curves", ACM Transactions on Computer Graphics, ACM SIGGRAPH 2007, San Diego, USA, 2007
- Martin Hachet, Fabrice Dècle, Sebastian Knödel, Pascal Guitton, "Navidget for Easy 3D Camera Positioning from 2D Inputs", Proceedings of IEEE 3DUI - Symposium on 3D User Interfaces - 2008

32 Développement d'un outil pour le doigté automatique de partitions de piano

Responsables : Pierre Hanna, Matthias Robine, Antoine Rollet

lieu : LaBRI

téléphone : 05.40.00.35.07

e-mail : pierre.hanna@labri.fr

équipe : Image et Son, thème : musique informatique

Mots-clés : doigté de piano, midi, modélisation de contraintes physiques, aide à la musique, musique informatique

Description du sujet

L'informatique permet de plus en plus d'assister les musiciens dans les phases de conception et de pratique de la musique. Ainsi, grâce à des interfaces adaptées, il est possible de saisir des données musicales, de les analyser, ou encore de les utiliser pour générer des sons.

L'objectif de ce stage est de développer une application d'aide au doigté de partitions pour piano. En effet, il existe une grande quantité de morceaux disponibles dans des formats musicaux tels que le format MIDI, mais ces formats ne contiennent aucune information sur le choix du doigté. On constate qu'un pianiste professionnel va appliquer une démarche qui lui est naturelle pour trouver le doigté qui lui convient, mais cette étape peut être longue voire impossible pour un pianiste amateur.

Des travaux ont déjà eu lieu dans ce domaine, et ont permis notamment d'identifier les difficultés principales liées aux contraintes physiques de la main. Toutefois, ces travaux se limitent généralement à des morceaux monophoniques et d'une seule main.

Après avoir pris connaissance de la littérature dans ce domaine, le stagiaire proposera une modélisation du problème, et ensuite développera sur cette base une application prenant en entrée un fichier au format MIDI, et fournissant en sortie une partition (au format pdf par exemple) qui propose un doigté "optimal" pour le morceau, avec pour objectif d'améliorer les travaux précédents (deux mains, polyphonie).

Commentaire : Des connaissances en solfège et en pratique du piano sont nécessaires pour réaliser ce travail.

Références bibliographiques

- [PSCRD97] Richard Parncutt, John A. Sloboda, Eric F. Clarke, Matti Raekallio, and Peter Desain. "An ergonomic model of keyboard fingering for melodic fragments," *Music Perception*, vol. 14, no. 4, pp. 341-382, 1997.
- [Rob07] Mathias Robine. "Analyse automatique du doigté au piano". In *proceedings of Journées d'Informatique Musicale (JIM07)*, Lyon, France (2007)
- <http://www.rosegardenmusic.com/>
- <http://common-lisp.net/project/gsharp/>

33 Analyse musicale par tatouage audio-numérique

Responsable : Sylvain Marchand

lieu : LaBRI

téléphone : 05.40.00.38.74

e-mail : sylvain.marchand@labri.fr

équipe : Image et son, thème : Modélisation du son et de la musique, projet : ANR DReaM

Mots-clés : son, tatouage, stéganographie, informatique musicale

Description du sujet

Le tatouage (*watermarking*) audio-numérique consiste à cacher des informations à l'intérieur des sons, et ce de manière inaudible.

Nous proposons ici d'utiliser cette possibilité afin d'inclure dans le son un motif de tatouage suffisamment robuste pour résister aux transformations musicales élémentaires, tout en étant assez souple pour subir également ces transformations en même temps que le son support.

Au final, le but serait, en analysant le son après transformation, et plus précisément la version déformée du motif de tatouage, d'être capable de déterminer les déformations subies par le motif de tatouage initial et d'en déduire une chaîne probable de transformations ayant conduit au son actuel.

Ceci pourrait par exemple permettre d'analyser les pratiques des compositeurs.

On peut aussi envisager de retrouver le son d'origine dans le cas de transformations réversibles.

Il s'agit dans un premier temps de constituer une liste des techniques de tatouage existantes et des transformations sonores les plus utilisées, et dans un second temps déterminer quels sous-ensembles de ces techniques et de ces transformations permettent de répondre à la problématique posée.

Références bibliographiques

- Cléo Baras, Tatouage informé de signaux audio numériques Thèse de doctorat, Télécom Paris, juin 2006.

SAtANas

34 Ordonnancement structuré de tâches OpenMP

Responsable : Olivier Aumage

lieu : LaBRI/INRIA (A29 bis)

téléphone : 05.24.57.41.19

e-mail : olivier.aumage@inria.fr

équipe : SATANAS/RUNTIME, thème : Supports exécutifs

Mots-clés : threads, ordonnancement, support exécutif, multi-coeur, calcul intensif

Contexte

La transition récente de l'informatique vers les architectures multi-coeurs met en exergue la notion de *coût* du parallélisme : alimenter en travail des coeurs de plus en plus nombreux nécessite un parallélisme de plus en plus fin, à tel point que le temps de gestion du parallélisme n'est plus négligeable devant le temps de travail parallélisé.

En particulier la création d'un thread pour chaque activité potentiellement parallèle devient prohibitive et les techniques popularisées par le langage Cilk consistant à mutualiser les threads pour l'exécution de *tâches* élémentaires prennent désormais tout leur sens. C'est d'ailleurs la raison première de l'apparition ces dernières années d'outils tels que les Threading Building Blocks d'INTEL et de l'évolution de la spécification OpenMP vers le support des tâches.

Problématique

Approche gestionnaire du parallélisme : Fondamentalement, la difficulté à surmonter est la nécessité d'un bon équilibre dynamique entre la *production* de parallélisme (la génération de tâches) et la *consommation* de parallélisme (l'exécution de tâches). Si l'équilibre penche en faveur de la production de parallélisme, l'allocation de ressources pour stocker les tâches en attente d'exécution risque d'être sous-optimale. À l'inverse, si l'on favorise trop la consommation on augmente le risque de déséquilibres de charge et de temps morts des coeurs.

Intégration des acquis ; Par ailleurs, le facteur coût du parallélisme ne se substitue pas aux autres facteurs influençant l'efficacité des applications parallèles mais il s'y ajoute. Des aspects tels que la localité et l'affinité, par exemple, désormais bien connus (si ce n'est maîtrisés...) dans le parallélisme de threads, restent plus que jamais critiques à l'échelle de tâches élémentaires. Ces aspects devraient donc idéalement être pris en compte par le support exécutif d'ordonnancement lors des trois phases essentielles de la vie d'une tâche : la génération, l'exécution et le vol.

Articulation : Enfin, l'articulation entre le niveau d'ordonnancement coopératif de tâches et le niveau d'ordonnancement préemptif de threads (lui-même multiple avec le niveau threads utilisateurs, le niveau threads noyaux/processus, voire même le niveau OS en cas de virtualisation... !) est à la fois source de flexibilité et de questionnement quant-à la recherche de performance et d'optimalité

Travail demandé

En s'appuyant sur l'expérience acquise dans la réalisation des logiciels Marcel (threads), Bubblesched (ordonnancement hiérarchique) et ForestGOMP (support exécutif OpenMP), il s'agit d'étudier cette problématique selon les pistes (non exhaustives) suivantes :

- en définissant les modalités d'intégration du support des tâches OpenMP dans ForestGOMP,

- en explorant l'évolution en conséquences des algorithmes originaux d'ordonnancement de ForestGOMP tels que Cache ou Memory,
- en caractérisant les nécessaires sources d'information permettant de guider le support exécutif dans son cheminement.

Pointeurs

- <http://runtime.bordeaux.inria.fr/>
- <http://runtime.bordeaux.inria.fr/forestgomp/>
- <http://runtime.bordeaux.inria.fr/marcel/>
- <http://www.openmp.org/>

Références bibliographiques

- The Design of OpenMP Tasks. Eduard Ayguadé and Nawal Copty and Alejandro Duran and Jay Hoeflinger and Yuan Lin and Federico Massaioli and Xavier Teruel and Priya Unnikrishnan and Guansong Zhang. IEEE TPDS Vol. 20 Num. 3 pages 404–418.
- The Implementation of the Cilk-5 Multithreaded Language. Frigo M., LEISERSON C. E., RANDALL K. H. ACM SIGPLAN '98 Conference on Programming Language Design and Implementation, Montreal, 1998. Pages 212–223.

35 Modèle de performance pour architectures multicoeurs hiérarchiques

Responsable : Denis Barthou et Brice Goglin

lieu : LaBRI

téléphone : 05.24.57.41.16

e-mail : `Denis.Barthou@inria.fr`

équipe : Satanas, thème : Runtime

Mots-clés : Multicoeurs, NUMA, hiérarchique, modèle de performance

Description du sujet

L'augmentation du nombre de coeurs dans les architectures récentes traduit un changement majeur et durable pour les machines de calcul qui contiennent désormais différents niveaux de caches partagés entre certains coeurs, plusieurs sockets et/ou noeuds mémoire NUMA. Les affinités entre tâches (communications, synchronisations, partage de données, etc.) et entre tâches et données deviennent un critère majeur de placement, d'ordonnancement et d'optimisation sur de telles architectures, que ce soit pour des applications parallèles de type MPI ou OpempMP, voire même des applications séquentielles.

Pour optimiser son codes hautes performances, le programmeur doit faire face à un problème de passage à l'échelle posé par de telles architectures : le même code n'aura probablement ni la même structure ni la même stratégie de placement si l'architecture cible est différente. La modélisation des performances reste l'approche la plus adaptée pour guider l'optimisation parallèle. L'analyse et la modélisation des performances pour l'optimisation du placement des tâches MPI sont un champ très actif de recherche. Une approche plus globale, prenant en compte le parallélisme MPI et OpenMP pour aider à reformuler ce parallélisme sur de nouvelles architectures reste toutefois à faire.

L'objectif de ce mémoire est la modélisation des performances d'applications parallèles sur des architectures multicœurs afin de guider l'optimisation de code parallèle, notamment l'ordonnancement et le placement, en collaboration avec les autres travaux de l'équipe Runtime. Il s'agira donc dans un premier temps de proposer un certain nombre de microbenchmarks permettant par exemple d'évaluer le coût concret d'un partage de données dans les différents niveaux de cache, ou d'accès mémoire depuis différents sockets ou noeuds NUMA. Par la suite, ces travaux pourront servir à l'implémentation et l'optimisation d'une barrière hiérarchique entre threads en tenant compte de la topologie de la machine. Plus généralement, ces informations quantitatives auront pour objectif de fournir d'une part des critères d'ordonnancement pour les bibliothèques multithreads Marcel ou OpenMP Forest-GOMP, et d'autre part du feedback pour les développeurs de codes parallèles.

La valorisation de l'approche et du modèle de performances qui en découle se fera ensuite par leur intégration dans l'outil MAQAO, développé en collaboration avec le laboratoire EXATEC (Intel, UVSQ, CEA, Genci). MAQAO est un outil d'analyse statique et dynamique de code et de mise au point de performances.

Ces travaux pourront notamment être réalisés sur des machines parallèles à 8, 16, 24 puis 96 cœurs à mémoire partagée dont dispose l'équipe Runtime.

Ce sujet de mémoire pourra éventuellement donner lieu à une poursuite en thèse.

<http://www.prism.uvsq.fr/~bad/Research/> <http://runtime.bordeaux.inria.fr/>

Références bibliographiques

- François Broquedis, Nathalie Furmento, Brice Goglin, Raymond Namyst, and Pierre-André Wacrenier. Dynamic Task and Data Placement over NUMA Architectures : an OpenMP Runtime Perspective. In *Evolving OpenMP in an Age of Extreme Parallelism*, 5th International Workshop on OpenMP, IWOMP 2009, volume 5568 of *Lecture Notes in Computer Science*, Dresden, Germany, pages 79-92, June 2009. Springer.
- Samuel Thibault, Raymond Namyst, and Pierre-André Wacrenier. Building Portable Thread Schedulers for Hierarchical Multiprocessors : the BubbleSched Framework. In *EuroPar*, Rennes, France, 8 2007. ACM.

- Denis Barthou, Andres Charif-Rubial, William Jalby, Souad Koliai, and Cedric Valensi. Performance tuning of x86 openmp codes with maqao. In Parallel Tools Workshop, Dreden, September 2009. Lect. Notes in Computer Science. To appear.
- A. Ketterlin and P. Clauss. Prediction and trace compression of data access addresses through nested loops recognition. In IEEE/ACM International Symposium on Code Generation and Optimization, CGO. IEEE/ACM, april 2008.
- Guillaume Mercier and Jérôme Clet-Ortega. Towards an efficient process placement policy for mpi applications in multicore environments. In EuroPVM/MPI, volume 5759 of Lecture Notes in Computer Science, pages 104115, Espoo, Finland, September 2009. Springer.
- Jaydeep Marathe, Frank Mueller, Tushar Mohan, Sally A. McKee, Bronis R. de Supinski, and Andy Yoo. Metric : Memory tracing via dynamic binary rewriting to identify cache inefficiencies. ACM Trans. Program. Lang. Syst., 29(2), 2007.

36 Algorithme hiérarchique parallèle pour l'évaluation sur une grille d'un champ électrique

Responsable : Olivier Coulaud

lieu : INRIA

téléphone : 05 24 57 40 80

e-mail : `Olivier.Coulaud@inria.fr`

équipe : SATANAS , thème : Algorithmique et applications hautes performances, projet : HIEPACS

Mots-clés : methode hiérachique, octree, methode multipoles rapide

Description du sujet

La modélisation des matériaux hybrides constitués de molécules organiques incluses dans une structure inorganique (argile, silice, etc.) sont un enjeu majeur dans de nombreux domaines comme pour les peintures, la dépollution de l'air, etc. L'étude des propriétés dynamiques de ces matériaux hybrides, par exemple les spectres électroniques, etc. nécessite de coupler un modèle de mécanique quantique pour calculer les états excités avec un modèle de dynamique moléculaire pour faire évoluer le système dans le temps. On parle de couplage mécanique quantique (QM) / mécanique moléculaire (MM). Ce couplage est réalisé au travers de deux codes parallèles : le code SIESTA pour la chimie quantique et le code DL_POLY pour la

partie moléculaire. L'environnement est traité par la mécanique moléculaire tandis que la molécule organique est traitée par la méthode quantique.

Pour calculer l'influence de l'environnement (la structure cristalline) sur la molécule organique (le colorant), nous devons évaluer le champ électrique induit par les particules de la structure inorganique sur une grille de points englobant le colorant. Dans l'exemple de la polygorskite, l'environnement est constitué d'environ 400 000 atomes et la grille d'environ un million de points. Une approche directe (double boucle) ne permet pas de calculer rapidement ce champ et conduit à des temps de simulation prohibitif. Pour diminuer cette complexité, nous introduisons une approche hiérarchique basée sur la méthode des multipôles. Il s'agit de découper l'espace de simulation dans un octree et d'approcher le champ lointain par une approximation polynomiale.

Travail à réaliser

Il s'agit de développer un algorithme rapide et parallèle pour évaluer le champ électrique dû aux particules sur une grille de points par la méthode des multipôles. Pour réaliser cet algorithme on adaptera le code FMB à notre problème. A partir de la distribution des atomes dans DL_POLY, il s'agira dans un premier temps de construire une distribution spatiale des atomes basée sur la construction d'un octree puis de calculer à l'aide de la bibliothèque FMB le calcul du champ sur la grille ayant une décomposition bloc-cyclique en Y et Z. Enfin, au cours des itérations en temps, il faut maintenir à jour la distribution des atomes dans l'octree.

L'adaptation du code FMB [1] se fera dans deux directions. La première consistera à séparer les sources (points participant à la création du champ) des cibles (points où l'on évalue le champ). La deuxième modification portera sur la prise en compte des conditions aux limites. Cette approche nous permet de considérer un cristal de taille "infinie" et non pas la cellule élémentaire pour évaluer le champ électrique sur le maillage. Pour valider la méthode on l'implémentera dans le couplage QM/MM avec les codes SIESTA et DL_POLY soit directement dans le code DL_POLY soit en introduisant un troisième code dans le couplage qui sera en charge d'évaluer ce potentiel.

Le stage se déroulera dans l'équipe projet INRIA HiePACS .

Références bibliographiques

- 1 P. Fortin, thèse, Algorithmique hiérarchique parallèle haute performance pour les problèmes à N-corps <http://tel.archives-ouvertes.fr/tel-00135843>.
- 2 C. Lambert, T. Darden and J. Board, J. Comp. Phys. 126 (1996).
- 3 Matt Challacombe, Chris White, and Martin Head-Gordon, Periodic boundary conditions and the fast multipole method. J Chem. Phys., 107 (23) 1997.

37 Optimisations inter-processus des communications réseaux

Responsable : Alexandre Denis

lieu : INRIA/LaBRI

téléphone : 05.24.57.41.14

e-mail : `Alexandre.Denis@labri.fr`

équipe : Satanas, thème : Runtime, projet : Runtime

Mots-clés : communications réseau, grappe, modèle de programmation hybride.

Description du sujet

Depuis quelques années, l'émergence des processeurs multi-coeurs a radicalement changé la physionomie des machines courantes, ce qui a eu des repercussions sur les grappes de calcul. Alors qu'il était typique auparavant qu'une grappe soit constituée de noeuds à deux processeurs, la tendance actuelle se porte plutôt vers des noeuds dotés de deux ou quatre processeurs, dotés chacun de deux à quatre coeurs, pour un total jusqu'à seize coeurs par noeud. Cette tendance à l'augmentation du nombre de coeurs par noeud est nette et semble durable. Cette évolution du matériel a des repercussions à plusieurs niveaux, en particulier sur le modèle de programmation et sur l'accès au réseau.

Avec l'augmentation du nombre de coeurs par noeud, le modèle de programmation qui était standard jusque là, à savoir une programmation du parallélisme au travers de la seule interface MPI (Message Passing Interface) avec un processus de calcul par coeur, n'est plus adapté. Il conduirait à de très nombreux processus par noeud, entraînant une importante consommation mémoire et un échange de nombreux messages MPI au sein d'un noeud, conduisant à des performances suboptimales car ces messages pourraient être évités étant donné que les processus tournent sur le même noeud donc pourraient partager directement de la mémoire. Pour résoudre ce problème de passage à l'échelle de MPI, une solution qui commence à être assez largement acceptée dans la communauté consiste à utiliser le langage OpenMP pour gérer le parallélisme en mémoire partagé à l'intérieur d'un noeud, et de cantonner MPI au passage de messages inter-noeud, ce qui mène à une solution hybride OpenMP+MPI.

Cependant, cette solution a des repercussions au niveau de l'accès au réseau. En effet, les implémentations MPI en particulier, et les bibliothèques d'accès au réseau d'une manière générale, ne sont pas conçues pour fonctionner en environnement multi-thread. Il se pose de nombreux problèmes tels qu'une concurrence entre les threads de calcul et les threads réalisant des accès au réseau menant à un problème de réactivité, ou encore des accès par plusieurs threads simultanément et de façon

concurrente à une même carte réseau. Ces questions sont l'objet de la thèse de François Trahay, soutenue en décembre 2009.

L'objet de ce sujet de mémoire est de s'intéresser aux problématiques posées à bas niveau dans les bibliothèques de communication, consécutivement aux évolutions des architectures et des modèles de programmation hybrides. Il s'agira alors pour le candidat de s'intéresser aux aspects suivants :

- Imaginer un canevas capable de gérer à bas niveau plusieurs cartes réseau, accédées simultanément par plusieurs processus, chaque processus étant lui-même composé de plusieurs threads. Il semble opportun d'envisager une optimisation et un ordonnancement des messages à un niveau inter-processus, globalement sur chaque noeud. Par exemple, notre implémentation actuelle NewMadeleine est capable d'agréger des messages consécutifs d'une application en un seul paquet émis sur le réseau ; il est attendu d'appliquer des schémas d'optimisation similaires mais à un niveau global entre plusieurs processus.
- Tirer profit des nouvelles opportunités d'optimisation offertes par l'augmentation du nombre de ressources et du nombre de processus. Par exemple, nos stratégies actuelles de multi-rail gèrent deux cartes réseaux et découpent systématiquement les paquets pour maximiser le taux d'occupation des cartes. Avec des cartes réseaux plus nombreuses et plusieurs processus, il devient pertinent d'envisager une adaptation dynamique du nombre de cartes allouées aux communications pour garder de la bande passante disponible pour les éventuelles communications qui pourraient être issues d'autres threads ou processus.
- Étudier le passage à l'échelle des mécanismes impliqués, aussi bien en nombre de threads qu'en nombre de processeurs et de cartes réseau. Il semble désormais acquis qu'il n'est plus approprié de ne considérer que des cas particuliers avec un ou deux processeurs et une ou des cartes réseau, mais qu'il faille au contraire considérer N ressources.
- Valider la proposition par une implémentation, évaluer les performances par des micro-benchmarks, et évaluer le comportement global de la solution proposée avec des applications complètes. À cette occasion, l'étudiant sera amené à travailler en collaboration avec les autres membres de l'équipe s'intéressant aux modèles de programmation hybrides.

<http://runtime.bordeaux.inria.fr/>

Références bibliographiques

- François Trahay. De l'interaction des communications et de l'ordonnancement de threads au sein des grappes de machines multi-coeurs. PhD thesis, Université Bordeaux 1, November 2009.
- François Trahay and Alexandre Denis. A scalable and generic task schedu-

- ling system for communication libraries. In Proceedings of the IEEE International Conference on Cluster Computing, New Orleans, LA, September 2009. IEEE Computer Society Press.
- François Trahay, Élisabeth Brunet, Alexandre Denis, and Raymond Namyst. A multithreaded communication engine for multicore architectures. In CAC 2008 : Workshop on Communication Architecture for Clusters, held in conjunction with IPDPS 2008, Miami, FL, April 2008. IEEE Computer Society Press.

38 Algorithmique de l'équilibrage de charge pour des couplages de codes complexes

Responsables : Aurélien Esnard ; Olivier Coulaud ; Jean Roman

lieu : INRIA, Bat. A29 bis, Université Bordeaux 1

téléphone : 05.40.00.41.08 / 40.80 / 40.33

e-mail : aurelien.esnard@labri.fr, olivier.coulaud@inria.fr,
jean.roman@labri.fr

équipe : SAtANas, thème : HiePACS

Mots-clés : parallélisme, simulation numérique, couplage de codes, partitionnement, graphe, hypergraphe, MPI.

Description du sujet

Dans le contexte du calcul scientifique haute performance (ou HPC), l'équilibrage de la charge est un problème crucial pour répartir le travail entre les différents processus d'une simulation numérique parallèle, afin de minimiser le temps global d'exécution. Ce problème étant déjà bien étudié, la stratégie la plus répandue consiste à modéliser la simulation à l'aide d'un graphe et à découper ce graphe en parties de poids équilibrées avec un outil de partitionnement de graphes comme SCOTCH ou METIS. À titre d'exemple, considérons une simulation en mécanique des fluides qui utilise un maillage à base d'éléments triangulaires comme support du calcul. Afin d'équilibrer la charge en vue d'une exécution parallèle sur 4 processeurs, on peut construire un graphe en associant un sommet à chaque élément du maillage, deux sommets étant connectés dans le graphe si les éléments sont adjacents dans le maillage. En partitionnant ce graphe en 4 parties, on en déduit un découpage du maillage, utilisé pour répartir les calculs.

De nos jours, les simulations numériques tendent à se complexifier, notamment en mixant plusieurs codes représentant des physiques différentes ou des échelles différentes. On parle alors de couplage de codes multi-physiques ou multi-échelles.

L'exemple le plus fameux dans ce domaine est très certainement le cas des simulations en climatologie qui interconnectent de multiples codes modélisant les océans, l'atmosphère, les glaces aux pôles, etc. Dans ce contexte, le problème de l'équilibrage de charge devient également plus difficile, car il ne s'agit plus d'équilibrer chacun des codes séparément, mais l'ensemble de ces codes pris dans leur globalité et en particulier au niveau des processeurs travaillant dans la zone de couplage. Par ailleurs, la nature des codes couplés (structure de données, calculs associés) pouvant être très différente, il est important de considérer ce problème de manière générique par une modélisation appropriée : le graphe ou l'hypergraphe de couplage.

Dans l'équipe SAtANas-HiePACS, nous nous intéressons au développement de nouvelles méthodes algorithmiques pour l'équilibrage des codes couplés. Une première stratégie d'équilibrage a été mise en oeuvre dans la bibliothèque LBC2 (Load Balancing for Code Coupling). Cette stratégie donne des résultats encourageants pour des cas simples en 2D, mais il faut généraliser et valider cette méthode pour des géométries plus complexes en 2D et 3D. D'autres stratégies d'équilibrage pourront être imaginées. Au final, ces travaux seront intégrés dans une "vraie" simulation multi-échelle en physique des matériaux, couplant un modèle atomique et un modèle continu d'élasticité. Des expérimentations en vraie grandeur pourront être menées sur une plate-forme de calcul haute-performance disposant d'un grand nombre de coeurs de calcul.

Références bibliographiques

- Scotch
- Metis
- Zotan
- PaToH
- Mondrian

39 Placement dynamique d'applications sur machine non-uniformes multicoeur

Responsables : Emmanuel Jeannot et Guillaume Mercier

lieu : LaBRI

téléphone : 05.24.57.40.93

e-mail : Emmanuel.Jeannot@labri.fr

équipe : Supports et AlgoriThmes pour les Applications Numériques hAutes performanceS, thème : Les supports exécutifs hautes performances, projet : Runtime

Mots-clés : Calcul parallèle, placement, machine NUMA.

Description du sujet

Dans de nombreux domaines scientifiques et d'ingénierie, il est nécessaire de résoudre des problèmes complexes réclamant d'énormes besoins en calculs. Dans ce but, on a utilisé des ordinateurs parallèles homogènes. Cependant, les récentes avancées en architectures font que les calculateurs haute performance sont de plus en plus hiérarchiques et hétérogènes (ex : une grappe de nœuds multiprocesseurs avec plusieurs cœurs, interconnectés avec différentes technologies de réseau haut débit). À l'avenir les machines parallèles auront couramment des milliers de nœuds composés de processeurs de plusieurs dizaines de cœurs. Du fait de l'hétérogénéité et de la hiérarchie des infrastructures, programmer efficacement et simplement de tels environnements est un des défis majeur du calcul haute performance.

En particulier, le placement des processus de calcul sur de telles machines joue un rôle crucial sur les performances de l'application du fait de l'affinité de certains processus entre eux et des performances inhomogène des communications. Or, le placement par défaut proposé par des bibliothèques de calcul parallèle comme MPI (<http://www.mcs.anl.gov/research/projects/mpi>) ou CHARM++ (<http://charm.cs.uiuc.edu/research/>) est certes bien adapté pour les architectures à plat (sans hiérarchie), mais pose des problèmes important dans le cas hiérarchique car il ne tient pas compte de l'affinité des processus entre eux.

De plus, la manière dont sont échangés les messages évolue au cours de l'exécution de l'application. Ainsi, un bon placement initial des processus peut s'avérer très inefficace en fin d'exécution. Il faut donc être capable de le changer dynamiquement en s'adaptant aux conditions d'exécution.

L'objectif de ce stage est de travailler sur les aspects du placement dynamique des processus pour les applications haute-performance. Au sein de l'équipe Santanas du LaBRI et en particulier de l'EPI INRIA Runtime, de nombreux outils et techniques existent ou sont en cours de développement qui contribuent à la mise en œuvre de cet objectif. C'est le cas de la bibliothèque HWLOC (<http://www.open-mpi.org/projects/hwloc/>) qui permet d'obtenir la topologie d'un environnement cible. D'autre part, nous avons développé des algorithmes de placement statique de processus. Enfin, plusieurs collaborations internationales et en particulier avec les USA (Univ. Urbana Champaign, Illinois) nous fournissent de nombreux cas concrets pour la mise en œuvre des solutions que nous pourrions proposer.

Le sujet décrit ici vise à la mise en œuvre de stratégies efficaces de placement dynamique de processus. Nous proposons le plan de travail suivant :

1. **Étude** : À partir du placement par défaut de MPI ou de CHARM++, cette

première partie a pour but d'étudier et de mettre au point des outils et des stratégies dynamiques permettant, à l'exécution, d'adapter le placement pour prendre en compte les imperfections du placement initial et les évolutions de l'application. Il s'agit aussi de prendre en compte les goulets d'étranglement de l'architecture et des communications entre processus. Pour cela, on étudiera des mécanismes permettant, à l'exécution, d'évaluer les insuffisances du placement courant.

2. **Mise en œuvre** : L'objectif est d'implanter les solutions proposées et de les tester sur des applications réelles. En particulier, dans le cadre de notre collaboration avec l'Université d'Urbana Champaign (Illinois USA), nous souhaitons implanter ces mécanismes pour améliorer l'exécution l'application NAMD (dynamique moléculaire : <http://www.ks.uiuc.edu/Research/namd>), développée en CHARM++. Cette dernière partie pourra conduire à une visite à l'université d'Urbana Champaign.

Références bibliographiques

- Guillaume Mercier and Jerome Clet-Ortega, *Towards an Efficient Process Placement Policy for MPI Applications in Multicore Environments*, in Recent Advances in Parallel Virtual Machine and Message Passing Interface , Proceedings of the 16th European PVM/MPI users' group Meeting, Volume 5749 of Lecture Notes in Computer Science, pages 104-155, Espoo, Finland, September 2009. Springer.
- Hu Chen, Wenguang Chen, Jian Huang, Bob Robert, H. Kuhn, *MPIPP : an automatic profile-guided parallel process placement toolset for SMP clusters and multiclusters* Proceedings of the 20th annual international conference on Supercomputing table of contents Pages : 353–360
- Jesper Larsson Träff, *Implementing the MPI process topology mechanism* Proceedings of the 2002 ACM/IEEE conference on Supercomputing table of contents Baltimore, Maryland Pages : 1 – 14, 2002.
- Laxmikant V. Kale, Eric Bohm, Celso L. Mendes, Terry Wilmarth and Gengbin Zheng, *Programming Petascale Applications with Charm++ and AMPI*, in Petascale Computing : Algorithms and Applications, CRC Press
- James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeeln Laxmikant Kalé and Klaus Schulten *Scalable Molecular Dynamics with NAMD* J. Comput Chem. 2005 December ; 26(16) : 1781–1802.

40 Modèles d'exécutions hybrides pour architectures hétérogènes

Responsable : Raymond Namyst

lieu : LaBRI, INRIA

téléphone : 05 24 57 40 34

e-mail : Raymond.Namyst@labri.fr

équipe : Satanas, thème : Runtime

Mots-clés : Parallélisme, multicoeur, OpenMP, accélérateur, GPU.

Description du sujet

Le sujet de recherche proposé s'inscrit dans le cadre des travaux menés au sein de l'équipe INRIA RUNTIME qui s'intéresse à la conception de supports exécutifs performants pour architectures parallèles.

L'évolution des architectures parallèles contemporaines, qui intègrent de plus en plus des technologies accélératrices telles que les GPU, pose de nouveaux problèmes en raison du caractère hétérogène de leurs unités de calcul. De nombreuses équipes de recherche travaillent à la conception de modèles de programmation uniformes qui faciliteraient l'exploitation de ces machines. C'est le cas par exemple de l'approche suivie par les concepteurs du langage StarSs (UPC, Barcelone), qui propose des directives «à la OpenMP» pour générer automatiquement des dépendances entre les tâches d'un programme, ces tâches pouvant s'exécuter sur CPU ou sur GPU. Toutefois, cette approche impose de programmer l'intégralité de l'application sous forme d'un graphe de tâches, ce qui n'est pas optimal dans beaucoup de situations.

L'idée au coeur de ce travail est de permettre d'exécuter des applications selon un modèle d'exécution hybride permettant la coexistence de parallélisme «multithread» typiquement issu d'OpenMP avec du parallélisme «graphe de tâches» issu d'un environnement tel que starSs. Il s'agit plus précisément d'étudier comment intégrer et articuler deux supports exécutifs développés dans notre équipe : StarPU (capable d'exécuter efficacement des graphes de tâches sur un ensemble de CPU+GPU) et ForestGOMP (un environnement d'exécution OpenMP permettant de maîtriser finement l'ordonnancement des threads sur architectures multicoeurs).

Le point central du travail sera d'évaluer, à l'aide d'applications synthétiques et d'applications réelles, les avantages et inconvénients respectifs des deux modèles d'exécution suivants : un modèle d'exécution OpenMP dans lequel des sous-graphes de tâches peuvent-être générés et partiellement exécutés sur des accélérateurs, et un modèle d'exécution centré sur l'exécution de graphes de tâches dans lequel certaines tâches sont parallélisées à l'aide d'OpenMP..

Une fois déterminé le meilleur compromis, une adaptation de l'environnement StarSs au-dessus de StarPU+ForestGOMP pourra être réalisée.

Références bibliographiques

- Eduard Ayguade, Rosa M. Badia, Daniel Cabrera, Alejandro Duran, Marc Gonzalez, Francisco Igual, Daniel Jimenez, Jesus Labarta, Xavier Martorell, Rafael Mayo, Josep M. Perez, and Enrique S. Quintana-Orti. A proposal to extend the openmp tasking model for heterogeneous architectures. In IWOMP'09 : Proceedings of the 5th International Workshop on OpenMP, pages 154-167, Berlin, Heidelberg, 2009. Springer-Verlag.
- Eduard Ayguadé, Rosa M. Badia, Francisco D. Igual, Jesus Labarta, Rafael Mayo, and Enrique S. Quintana-Orti. An Extension of the StarSs Programming Model for Platforms with Multiple GPUs. In Proceedings of the 15th Euro-Par Conference, Delft, The Netherlands, August 2009.
- François Broquedis. Ordonnancement de threads OpenMP et placement de données coordonnés sur architectures hiérarchiques. In 19ème Rencontres Francophones du Parallélisme, Toulouse / France, September 2009.
- Cédric Augonnet. StarPU : un support exécutif unifié pour les architectures multicœurs hétérogènes. In 19ème Rencontres Francophones du Parallélisme, Toulouse / France, September 2009.

41 Mise oeuvre d'un schéma AMG pour le partitionnement multi-niveaux de graphes

Responsables : François Pellegrini (LaBRI & INRIA) et Ilya Safro (Argonne & Weizmann Institute)

lieu : LaBRI

téléphone : 05 24 57 40 61

e-mail : pelegrin@labri.fr safro@mcs.anl.gov

équipe : Satanas

Mots-clés : graphe, partitionnement, schéma multi-niveaux, appariements multi-grilles algébriques

Description du sujet

De nombreux problèmes d'optimisation peuvent être modélisés sous la forme d'un problème de partitionnement sous contraintes d'un graphe représentant le problème initial à résoudre. Pour calculer en temps acceptable des partitions de bonne qualité de gros graphes (d'une taille supérieure au million de sommets), la plupart des outils de partitionnement actuels, tels que le logiciel Scotch [1] développé

à Bordeaux au sein de l'équipe INRIA Bacchus, reposent sur un schéma multi-niveaux. Dans ce schéma, le graphe à partitionner est successivement contracté en graphes de plus petites tailles, par fusion de sommets voisins, puis une partition initiale est calculée sur le graphe le plus petit ainsi obtenu (d'une taille de quelques centaines de sommets), après quoi cette partition est projetée et raffinée de proche en proche sur les graphes les plus fins jusqu'à obtenir une partition du graphe initial.

La qualité finale de la partition obtenue repose à la fois sur la qualité de la partition initiale, des algorithmes de raffinement, mais aussi de la méthode d'appariement utilisée pour contracter les graphes. Cette méthode doit absolument préserver leur structure topologique, faute de quoi la partition initiale calculée sur le graphe le plus petit ne représenterait aucunement une solution approchée du problème initial.

La majorité des méthodes de contraction utilisées à ce jour sont basées sur des appariements d'arêtes (edge matchings), servant à la construction de graphes quotients par regroupement des sommets liés par les arêtes choisies. Ces méthodes fonctionnent en fait selon un principe de « tout ou rien » : soit le voisin d'un sommet est celui relié par l'arête sélectionnée, et il est fusionné, soit il n'est pas impacté. Cette méthode est simple, mais susceptible d'introduire des artefacts topologiques selon la façon dont les arêtes sont sélectionnées. Dans une vision moins « binaire » du processus de contraction, on pourrait imaginer qu'un sommet puisse s'apparier avec plusieurs de ses voisins en même temps, avec une force pouvant varier continûment entre 0 (pas de lien) et 1 (lien très fort), permettant ainsi de « lisser » les artefacts d'appariement.

C'est par exemple le cas du schéma AMG (Algebraic Multi-Grid) développé par Ilya Safro dans le cadre de la renumérotation de graphes [2], et appliqué par Cédric Chevalier et lui au partitionnement de graphes dans le cadre d'un travail exploratoire [3].

Le but de ce projet est d'évaluer dans un cadre pratique les gains apportés par cette technique, en intégrant la méthode AMG au sein du logiciel Scotch. Il s'agira de mener une étude expérimentale poussée de la méthode, sur des familles de graphes aux propriétés diverses, pour voir en quoi elle peut être plus robuste que les méthodes traditionnelles d'appariement basées sur le choix de l'arête de plus fort poids (appariement Heavy Edge Matching).

L'ensemble des développements effectués sera intégré dans une nouvelle méthode multi-niveaux mise à la disposition des usagers de la bibliothèque libScotch.

En fonction de l'avancée des travaux, la parallélisation de cette méthode pourra être envisagée, en se basant sur les algorithmes parallèles multi-niveaux existant déjà au sein de la version parallèle de Scotch, appelée PT-Scotch.

Commentaires : Ce projet se déroulera au centre INRIA Bordeaux Sud-Ouest,

au sein de l'équipe Bacchus, en relation par courriel avec Ilya Safro. Une bonne connaissance de la programmation en langage C est absolument nécessaire.

Références

- 1 <http://www.labri.fr/~pelegrin/scotch/>
- 2 <http://www.mcs.anl.gov/~safro/papers/band.pdf>
- 3 http://www.mcs.anl.gov/~safro/papers/part_lion3.pdf

42 Solveur linéaire parallèle hybride hiérarchique

Responsable : J. Roman / L. Giraud en collaboration avec E. Agullo et A. Guermouche

lieu : LaBRI

téléphone : 05.24.57.40.33

e-mail : Jean.Roman@labri.fr, Luc.Giraud@inria.fr

équipe : SAANas , thème : HiePACS

Mots-clés : Calcul très haute performance, passage à l'échelle, machines pétafloppiques massivement multiprocesseurs et multi-coeurs, algèbre linéaire creuse et solveurs hybrides, simulation numérique en vraie grandeur.

Description du sujet

Les grands défis scientifiques, aussi bien académiques qu'industriels et rencontrés dans les simulations complexes (optimisation, calcul par éléments finis avec des maillages de plus en plus fins et volumineux), nécessitent la résolution de problèmes d'algèbre linéaire creuse de plus en plus volumineux et coûteux. Dans ce contexte, il apparaît que les méthodes hiérarchiques et/ou hybrides, au sens où elles combinent plusieurs méthodes de base comme des méthodes directes ou itératives ou multigrilles, offrent un compromis prometteur entre robustesse et souplesse tout étant très performantes. L'utilisation de solveurs directs sur des sous problèmes indépendants introduit un premier niveau de parallélisme naturel ; la mise en oeuvre de solveurs directs eux mêmes parallèles autorise l'exploitation d'une hiérarchie de tâches parallèles de granularités variables. Des travaux sur ce type de méthodes numériques parallèles ont déjà été effectués et ont permis de mettre en évidence le potentiel de telles approches. Le passage à l'échelle lié à une scalabilité importante des applications scientifiques sur des centaines de milliers, voir des millions, de coeurs de calcul est un vrai challenge. Compte tenu de la complexité de la mise en oeuvre, une algorithmique hiérarchique basée sur l'utilisation de bibliothèques numériques ultra performantes et combinant plusieurs niveaux de parallélisme avec des grains de calcul et des mécanismes sous jacents appropriés (grain grossier avec MPI, grain moyen avec du multi-threading en mémoire partagée, grain fin avec

une grand nombre de threads sur des accélérateurs de calcul comme les GPUs) est indispensable.

C'est dans ce contexte que se situe l'objectif de ce stage, qui est de contribuer à la conception algorithmique et à la mise en oeuvre en vraie grandeur (milliard d'inconnues pour des problèmes 3D non structurés) d'un solveur haute performance d'algèbre linéaire creuse basé sur une approche hybride et hiérarchique (couplage de différentes méthodes directes, itératives en fonction des caractéristiques du système à résoudre et de l'architecture hétérogène du calculateur) et ayant vocation à utiliser au mieux ces futures machines. Dans ce contexte, on s'intéressera aussi plus particulièrement aux systèmes linéaires augmentés dont on cherchera à exploiter au mieux les structures. La validation expérimentale pourra se faire sur des cas tests issus de nos partenaires industriels comme le BRGM (imagerie sismique), le CEA/CESTA (électromagnétisme) et EDF (mécanique des structures).

Ce travail est dans la continuité des recherches menées depuis de nombreuses années au sein du projet INRIA Bordeaux - Sud-Ouest HiePACS qui est commun avec le LaBRI et se fera en collaboration avec le CERFACS à Toulouse. Ce sujet de PFE rémunéré se déroulera donc au sein de l'équipe projet INRIA HiePACS et pourra faire naturellement l'objet d'un Master Recherche. Des possibilités de poursuite de ce travail dans le cadre d'un doctorat sont tout à fait envisageables compte tenu de l'ampleur du sujet.

Références bibliographiques

- L. GIRAUD, A. HAIDAR, and L. T. WATSON. Parallel scalability study of hybrid preconditioners in three dimensions. *Parallel Computing*, 34 :363-379, 2008.
- L. GIRAUD and A. HAIDAR. Parallel algebraic hybrid solvers for large 3D convection- diffusion problems. *Numerical Algorithms, Numerical Algorithms*, 51(2) :151-177, 2009.
- P. HÉNON, P. RAMET, and J. ROMAN. On using an hybrid MPI-Thread programming for the implementation of a parallel sparse direct solver on a network of SMP nodes. In *Proceedings of Sixth International Conference on Parallel Processing and Applied Mathematics, Workshop HPC Linear Algebra*, Poznan, Pologne, LNCS 3911, pages 1050-1057, September 2005.

Langages, Systèmes et Réseaux

43 Découverte de service dans une flotte hautement volatile de terminaux mobiles communicants

Responsable : David Bromberg et Serge Chaumette

lieu : LaBRI

téléphone : 05.40.00.69.04

e-mail : david.bromberg@labri.fr serge.chaumette@labri.fr

équipe : LSR, thème : Muse, projet :

Mots-clés : flotte mobile, découverte de services

Description du sujet

Grâce aux progrès réalisés ces dernières années dans le domaine de la miniaturisation de composants électroniques, au faible coût des dispositifs informatiques ou objets communicants, et à l'émergence et l'expansion des technologies de réseaux sans fil basées ou non sur des infrastructures (WLAN, Bluetooth, GSM, GPRS, UMTS), l'Internet of Things devient une réalité.

Le réseau sous-jacent est uniquement formé par des liens deux à deux entre objets qui constituent ainsi un réseau ad hoc hautement instable. On se trouve alors face à des flottes d'objets/de terminaux mobiles communicants dont les caractéristiques majeures sont la forte volatilité et la forte mobilité.

Une des problématiques majeures est la conception d'un système hautement distribué capable de gérer les différents objets communicants d'un espace actif et en particulier la présentation (catalogue) des services qu'ils offrent.

Il n'existe pas aujourd'hui de proposition résistant à l'extrême volatilité du système (le problème central étant l'absence de serveur stable pouvant stocker les références/localisations des services disponibles).

Dans ce mémoire de recherche, on va donc se concentrer sur le référencement/la découverte de services. On adaptera/exploitera pour cela des algorithmes conçus dans le thème Muse, en particulier tout ce qui concerne le re-étiquetage des graphes dynamiques [1, <http://www.site.uottawa.ca/~casteig/files/these.pdf>] et en particulier la constitution de forêts couvrantes. Cet algorithme décrit dans [1], section 2.3.4 semble en effet un bon candidat, modulo une adaptation, à la gestion de la découverte de services dans le contexte considéré.

Cette étude s'accompagnera le cas échéant de la réalisation d'un prototype.

Références bibliographiques

- A.Casteigts, Contribution à l'Algorithmique Distribuée dans les Réseaux Mobiles Ad Hoc - Calculs Locaux et Réétiquetages de Graphes Dynamiques,

- PhD thesis, Université de Bordeaux, 2007
- A. Casteigts, S. Chaumette, and A. Ferreira. Characterizing Topological Assumptions of Distributed Algorithms in Dynamic Networks. In Proc. of 16th Intl. Conference on Structural Information and Communication Complexity (SIROCCO'09), volume 5869 of Lecture Notes in Computer Science, Piran, Slovenia, May 2009. Springer-Verlag.
 - Bryan, D.A., Lowekamp, B. B., Zangrilli, M., The Design of a versatile, secure P2PSIP communications architecture for the public internet, in IEEE international Parallel and Distributed Processing Symposium. April, 2008 : Lyon, France
 - Matuszewski, M., Kokkonen, E., Mobile P2PSIP - Peer-to-Peer SIP Communication in Mobile Communities, in 5th IEEE Consumer Communications and Networking Conference, 2008. CCNC 2008. Jan, 2008 : Las Vegas.
 - Patrick Stuedi, Marcel Bihr, Alain Remund, and Gustavo Alonso - SIPHoc : Efficient SIP Middleware for Ad Hoc Networks, in Middleware 2007, ACM/IFIP/USENIX 8th International Middleware Conference, Newport Beach, CA, USA, November 26-30, 2007

44 Programmation générative et Reconfiguration matérielle

Responsable : David Bromberg

lieu : LaBRI

téléphone : 05.40.00.28.69

e-mail : David.Bromberg@labri.fr

équipe : MUSE, thème : LSR, projet : Z2Z

Mots-clés : Middleware, Programmation générative, Re-programmation matérielle

Description du sujet

Ces dernières décennies, et a fortiori ces dernières années, ont été marquées par une dualité logiciel-matériel forte. Cette dualité est fondée sur constat simple : un logiciel ne peut pas fonctionner sans un matériel adéquat et vice versa. Implicitement, cela signifie aussi que les performances du logiciel seront améliorées lorsque le matériel utilisé pour son exécution sera amélioré, et réciproquement. Plus précisément, au regard des progrès réalisés dans le domaine de la miniaturisation de composants électroniques, et dans le domaine des technologies logicielles, le logiciel et le matériel sont de plus en plus imbriqués au sein de systèmes de plus en plus complexes. C'est le cas, par exemple, dans de très nombreux domaines tel

que l'automobile, l'aéronautique, le spatial, la domotique, l'informatique diffuse, le médical, la sécurité et/ou le militaire. L'étude de cette dualité logiciel-matériel révèle qu'il faut repenser la façon dont l'un et l'autre sont conçus.

L'objectif de ce stage de recherche est d'étudier, et d'analyser les concepts de programmation générative afin de concevoir et de promouvoir un cadre de développement matérielle génératif. Etant donné le choix d'un support matériel générique tel qu'un FPGA, cette démarche consiste à concevoir et à mettre en oeuvre, à partir d'une description de haut niveau du système matériel/logiciel souhaité, un générateur qui produira automatiquement le code du logiciel et celui du matériel associé (code permettant la reconfiguration/spécialisation du FPGA). Le travail réalisé s'appuiera sur les résultats du projet de recherche Z2Z.

<http://www.labri.fr/perso/bromberg/>

Références bibliographiques

- Yérom-David Bromberg, Laurent Réveillère, Julia L. Lawall, Gilles Muller : Automatic Generation of Network Protocol Gateways. Middleware 2009 : 21-41

45 Mise à jour épidémique et incrémentale des environnements ubiquitaires

Responsable : David Bromberg

lieu : LaBRI

téléphone : 05.40.00.28.69

e-mail : David.Bromberg@labri.fr

équipe : MUSE, thème : LSR, projet :

Mots-clés : Middleware, protocole épidémique, mise à jour incrémentale, plateforme mobile, Android

Description du sujet

Grâce aux progrès réalisés ces dernières années dans le domaine de la miniaturisation de composants électroniques, du faible coût des dispositifs informatiques ou objets communicants, de l'émergence et de l'expansion des technologies réseaux sans fil basées ou non sur des infrastructures (WLAN, Bluetooth, GSM, GPRS, UMTS), l'informatique ubiquitaire (IU) devient aujourd'hui une réalité. L'IU se définit comme un environnement saturé d'objets communicants susceptibles de coopérer de façon autonome et transparente afin d'améliorer l'interactivité et/ou l'expérience de l'utilisateur. Ces objets, venant de domaines aussi variés que l'électronique grand public, l'électroménager, les télécommunications, la domotique ou l'automobile, doivent tous pouvoir communiquer spontanément avec

d'autres objets de leur l'environnement immédiat. L'utilisateur se retrouve alors au centre d'un espace, appelé espace actif, qui fédère dynamiquement l'ensemble des objets communicants disséminés dans son environnement. L'infrastructure du réseau sous-jacent est uniquement formée par les objets communicants eux-mêmes qui forment un réseau ad hoc hautement volatile.

Une des problématiques majeures de l'IU est la conception d'un système hautement distribué capable de configurer, coordonner et gérer les différents objets communicants d'un espace actif, en tenant compte des spécificités du réseau ad hoc sous-jacents, hautement volatile et potentiellement à large échelle.

L'objectif de ce stage de recherche est d'étudier, d'analyser et d'étendre les protocoles épidémiques et/ou les protocoles utilisés par les gestionnaires de versions tel que svn, git, mercurial, bazaar, de façon à établir un mécanisme de mise à jour incrémentale des systèmes mobiles des différents objets communicants disséminés dans l'environnement de l'utilisateur. La plateforme de test sera une plateforme Android.

<http://www.labri.fr/perso/bromberg/>

Références bibliographiques

- Mirco Musolesi, Cecilia Mascolo, Stephen Hailes : EMMA : Epidemic Messaging Middleware for Ad hoc networks. *Personal and Ubiquitous Computing* 10(1) : 28-36 (2006)

46 Evaluation du modèle à composant SOFTIMA

Responsable : Charles Consel

lieu : LaBRI/INRIA

téléphone : 05.24.57.40.35

e-mail : charles.consel@labri.fr

équipe : Langages, Systèmes et Réseaux, thème : Technologie des langages pour les services de communication , projet : Phoenix

Mots-clés : Architecture logicielle, Modèles à composants, Etude comparative, Aéronautique

Description du sujet

Contexte : L'évolution des systèmes avioniques connaît une accélération au même titre que l'ensemble des solutions numériques du monde d'aujourd'hui. Faire face à cette réalité en assurant un niveau de compétitivité concurrentiel est un enjeu majeur pour le groupe Thales. Afin de répondre à cet enjeu, Thales a lancé le projet SoftIMA. Ce projet a pour objectif de fédérer l'ensemble des solutions et compétences nécessaires en un tout cohérent.

Pour cela, SoftIMA définit trois axes de développement complémentaires et à adresser en cohérence : architecture, ingénierie et plate-forme d'exécution. Pour lier ces trois axes de développement, Thales a développé le modèle à composant SoftIMA Component Model.

SoftIMA CM permet tout d'abord d'avoir une vue haut-niveau des systèmes logiciels grâce à son langage d'architecture logicielle (ADL). Cet ADL permet de décomposer un système logiciel en composants et de définir les connecteurs entre composants. Ces connecteurs décrivent des interactions entre ces composants. Le modèle à composant est associé à un compilateur. Ce compilateur génère du code dans lequel le développeur n'a plus qu'à implémenter sa logique applicative. Le compilateur génère également les exécutables permettant de déployer les composants sur la plate-forme d'exécution.

Objectif : L'objectif de ce stage est de comparer et d'évaluer le modèle à composant SoftIMA CM par rapport aux modèles à composants existants. Des modèles à composants industriels tels que Corba Component Model (CCM) ainsi que des modèles à composants académiques tels que Fractal ou DiaSpec seront utilisés comme éléments de comparaison avec SoftIMA CM.

Pour effectuer cette comparaison, le stagiaire étudiera tout d'abord un exemple représentatif du domaine de l'aéronautique développé avec SoftIMA CM. Il devra ensuite définir des critères d'évaluation de modèles à composants. Il adaptera ensuite l'exemple développé avec SoftIMA CM sur les autres modèles à composants. Pour cela, le stagiaire développera ce même exemple en utilisant les autres modèles à composants sélectionnés pour effectuer la comparaison. Enfin, il appliquera les critères d'évaluation sur les différents modèles à composants afin de pouvoir évaluer SoftIMA CM par rapport aux approches existantes.

En fonction de son avancée dans l'évaluation de SoftIMA CM, le stagiaire pourra être amené à intégrer dans SoftIMA CM des caractéristiques manquantes pour l'instant (par exemple : qualité de service, tolérance aux fautes, etc.).

Informations complémentaires : Ce stage est proposé par l'équipe-projet Phoenix en collaboration avec Thales Systèmes Aéroportés situé à Pessac.

<http://phoenix.labri.fr>

Références bibliographiques

- "A Generative Programming Approach to Developing Pervasive Computing Systems", D. Cassou, B. Bertran, N. Lorient and C. Consel, publié à GPCE 2009
- "Recursive and Dynamic Software Composition with Sharing", E. Bruneton, T. Coupaye and J.-B. Stefani, publié à ECOOP 2002
- Spécification du modèle à composant CCM : <http://www.omg.org/technology/documents/formal/compon>

47 Plateforme applicative de gestion des connexions d'un réseau recouvrant

Responsable : Damien Magoni

lieu : LaBRI

téléphone : 05.40.00.35.40

e-mail : <magoni>@labri.fr

équipe : LSR, thème : Comet, projet : CLOAK

Mots-clés : overlay, middleware, transport protocols

Description du sujet

L'objectif de ce travail est de définir une plateforme applicative permettant de gérer des connexions Internet de tous types de manière flexible. Une connexion est un point d'entrée/sortie pour le transfert bidirectionnel de bout en bout des données. Une connexion relie donc deux équipements quelconques connectés n'importe où dans l'Internet. La plateforme se situera entre les applications et le système d'exploitation d'un équipement donné. Elle offrira une nouvelle interface de programmation pour les applications, basée sur l'utilisation de noms d'entités et d'identifiants d'actions, qui permettra de créer des connexions dites applicatives. La plateforme transformera les appels applicatifs en appels systèmes classiques basés sur l'utilisation des adresses IP et des numéros de ports (i.e. sockets). La plateforme découplera ainsi l'application du système d'exploitation et donc de l'équipement. Cela permettra d'interrompre une connexion système utilisant un protocole de transport (e.g. TCP, SCTP, etc) puis de la rétablir avec d'autres paramètres (e.g. nouvelle adresse IP, nouveau protocole de transport, nouveau numéro de port, etc) sans interrompre la connexion applicative correspondante. Cette faculté permettra, au cours d'une même connexion applicative, de déplacer l'équipement entre différents réseaux de manière transparente (sans avoir à utiliser Mobile IP) ou bien encore de changer de protocole de transport selon les caractéristiques du réseau. De plus, la plateforme devra aussi permettre de transférer une connexion applicative d'une application donnée à une autre, si cela possible et désiré par l'utilisateur, là encore sans couper la connexion. Cette plateforme devra être définie en tenant compte des contraintes de l'architecture existante de l'Internet et des caractéristiques des systèmes d'exploitation. Elle sera implémentée sous la forme d'un intergiciel s'exécutant dans l'espace utilisateur du système hôte. Certains aspects de la plateforme (e.g. tel que la gestion des mémoires tampons des connexions applicatives) seront validés par simulation. La plateforme sera évaluée par l'implémentation d'un prototype.

<http://dept-info.labri.fr/~magoni/sujets>

Références bibliographiques

- Autonomous, Scalable, and Resilient Overlay Infrastructure. Khaldoon Shami, Damien Magoni, Pascal Lorenz. *Journal of Communications and Networks*, 8(4) :378-390, 2006.
- A Scalable Middleware for Creating and Managing Autonomous Overlays. Khaldoon Shami, Damien Magoni, Pascal Lorenz. *COMSWARE'07 - International Conference on Communication System Software and Middleware*, pp. 1-8, January 7-12, 2007, Bangalore, India.
- Scalable Distributed k-Resilient Name to Address Binding System for Overlays. Khaldoon Shami, Damien Magoni, Piotr Lipinski, Pascal Lorenz. *ICN'06 - International Conference on Networking*, pp. 46-54, April 23-29, 2006, Mauritius.
- Application Layer Addressing, Routing and Naming Framework for Overlays. Damien Magoni, Pascal Lorenz. *GLOBECOM'05 - IEEE Global Telecommunications Conference*, pp. 66-71, November 28-December 2, 2005, Saint Louis, Missouri, USA.

48 Système de nommage à indirections multiples

Responsable : Damien Magoni

lieu : LaBRI

téléphone : 05.40.00.35.40

e-mail : <magoni>@labri.fr

équipe : LSR, thème : Comet, projet : CLOAK

Mots-clés : overlay, middleware, DHT

Description du sujet

L'objectif de ce travail est de définir un système de nommage qui permettra, entre autres, de séparer le plan d'adressage IP de celui de l'identification des points de connexions TCP ou UDP. Cette séparation va nécessiter la création d'un nouvel espace de nommage au niveau de la couche session du modèle OSI ainsi que la définition d'une nouvelle interface de programmation des communications au niveau de cette même couche. Le nouvel espace de nommage contiendra des noms d'équipements (liés à des adresses IP), des noms d'entités (i.e. utilisateurs, liés à des équipements) et des noms de sessions (liés à des entités). Une application ne se servira plus d'une adresse IP et d'un numéro de port pour ouvrir une connexion mais elle se servira seulement du nom de l'entité de destination et d'un identifiant d'action. Ce système de nommage permettra, par exemple, de faciliter la mise en oeuvre transparente et simultanée de la mobilité et de la sécurité dans l'Internet en découplant les identifiants d'une communication des équipements sur lesquels elle

a lieu. Il faudra ensuite créer dans ce même système, des espaces de nommage applicatifs qui permettront via des résolutions de noms à indirections multiples d'implémenter de manière transparente et efficace des communications multipoints, des transferts de connexion, des changements d'utilisateurs, etc. Le système de nommage proposé devra supporter dynamiquement les changements de noms et il devra être réparti afin de pouvoir être extensible et donc de supporter le passage à l'échelle. Pour cela, il pourra se baser sur un modèle pair-à-pair et s'implémenter à l'aide de tables de hachage distribuées. Enfin, il devra être indépendant par rapport à tout système existant (e.g. DNS). Ce système devra être défini en tenant compte des contraintes de l'architecture existante de l'Internet, puis validé par simulation et enfin testé par l'implémentation d'un prototype.

<http://dept-info.labri.fr/~magoni/sujets>

Références bibliographiques

- Autonomous, Scalable, and Resilient Overlay Infrastructure. Khaldoon Shami, Damien Magoni, Pascal Lorenz. *Journal of Communications and Networks*, 8(4) :378-390, 2006.
- A Scalable Middleware for Creating and Managing Autonomous Overlays. Khaldoon Shami, Damien Magoni, Pascal Lorenz. *COMSWARE'07 - International Conference on Communication System Software and Middleware*, pp. 1-8, January 7-12, 2007, Bangalore, India.
- Scalable Distributed k-Resilient Name to Address Binding System for Overlays. Khaldoon Shami, Damien Magoni, Piotr Lipinski, Pascal Lorenz. *ICN'06 - International Conference on Networking*, pp. 46-54, April 23-29, 2006, Mauritius.
- Application Layer Addressing, Routing and Naming Framework for Overlays. Damien Magoni, Pascal Lorenz. *GLOBECOM'05 - IEEE Global Telecommunications Conference*, pp. 66-71, November 28-December 2, 2005, Saint Louis, Missouri, USA.

49 Approche langage pour l'adaptation dynamique des flux multimédia

Responsable : Daniel Négru et Laurent Réveillère

lieu : LaBRI

téléphone : 05.40.00.28.69

e-mail : {negru, reveillere}@labri.fr

équipe : LSR, thème : COMET/MUSE

Mots-clés : Langage métier, approche générative, intergiciel, flux multimédia, adaptation

Contexte

De nombreux dispositifs électroniques (téléphone, pda, assistant personnel, PC, settop-box, etc.) fournissent actuellement des capacités variées de rendu audio et vidéo et disposent de connectivité filaire et/ou sans fil. Toutefois, l'accès depuis ces terminaux nomades à des contenus multimédia nécessitent d'adapter les caractéristiques des flux audio et vidéo diffusés pour qu'ils soient 1) compatibles et 2) optimisés au terminal et à son environnement (bande passante par exemple). De plus, ces contraintes pouvant évoluer dynamiquement en fonction de la mobilité de l'utilisateur et de l'utilisation du terminal (autonomie par exemple), il convient d'adapter dynamiquement ces flux en fonction du contexte.

L'adaptation des flux multimédia à la source n'étant pas réaliste, dans un grand nombre de cas (e.g. diffusion), compte tenu de l'hétérogénéité des terminaux et des conditions d'utilisation, une approche prometteuse consiste à interposer au bon endroit entre l'émetteur et les récepteurs un dispositif capable d'assurer dynamiquement cette adaptation. Toutefois, le développement d'une telle passerelle est une tâche fastidieuse et source d'erreur, qui doit être répétée dès qu'un nouveau dispositif apparaît ou qu'un nouveau format émerge.

Objet du stage

L'objectif du stage proposé consiste à définir une approche novatrice permettant de spécifier à l'aide d'un langage métier de haut niveau les caractéristiques principales des différents formats (source et destinataires) et de générer automatiquement une implémentation efficace et robuste. Plus précisément, l'approche proposée vise à concevoir et développer un intergiciel métier pour supporter l'adaptation de flux multimédia, un langage pour décrire les politiques d'adaptation et une chaîne de compilation pour générer l'implémentation de ces politiques en reposant sur les primitives offertes par l'intergiciel.

50 Mise à jour épidémique et incrémentale des environnements ubiquitaires

Responsable : Laurent Réveillère David Bromberg

lieu : LaBRI

téléphone : 05.40.00.28.69

e-mail : Laurent.Reveillere@labri.fr

e-mail : David.Bromberg@labri.fr

équipe : MUSE, thème : LSR, projet :

Mots-clés : Middleware, application protocol heterogeneity, protocol specification, generative programming

Description du sujet

The emergence of networked devices in the home has made it possible to develop applications that control a variety of household functions. However, current devices communicate via a multitude of incompatible protocols, and thus gateways are needed to translate between them. Gateway construction, however, requires an intimate knowledge of the relevant protocols and a substantial understanding of low-level network programming, which can be a challenge for many application programmers. Recently, a new approach to gateway construction, *z2z*, has been proposed. This generative approach is based on a domain-specific language for describing protocol behaviors, message structures, and the gateway logic. *Z2z* includes a compiler that checks essential correctness properties and produces efficient code. It has been used to develop a number of gateways involving a range of issues common to protocols used in the home.

However, one has to provide a complete set of *z2z* specification in order to generate a fully operational gateway. Therefore, the number of specifications to write increases if the number of protocols to support increases.

The aim of this project is to define a specification language enabling one to formally specify the behavior of an application-level protocol. From the formal specifications of two protocols, it becomes possible to generate the complete *z2z* specification enabling the generation of a protocol gateway.

<http://www.labri.fr/perso/reveille/> <http://www.labri.fr/perso/bromberg/>

Références bibliographiques

- D. Bromberg, L. Réveillère, J. Lawall, and G. Muller, Automatic Generation of Network Protocol Gateways, in *Middleware 2009 10th International Middleware Conference ACM/IFIP/USENIX*, pp. 1-20, 2009.
- Verifying distributed systems : the operational approach. Tom Ridge. In *POPL 2009*

51 Automatic generation of gateways between network protocols and libraries

Responsable : Laurent Réveillère

lieu : LaBRI

téléphone : 05.40.00.28.69

e-mail : reveillere@labri.fr

équipe : LSR, thème : MUSE

Mots-clés : Langage métier, approche générative, intergiciel, protocoles

Contexte

The emergence of networked devices in the home has made it possible to develop applications that control a variety of household functions. However, current devices communicate via a multitude of incompatible protocols, and thus gateways are needed to translate between them. Gateway construction, however, requires an intimate knowledge of the relevant protocols and a substantial understanding of low-level network programming, which can be a challenge for many application programmers. Recently, a new approach to gateway construction, *z2z*, has been proposed. This generative approach is based on a domain-specific language for describing protocol behaviors, message structures, and the gateway logic. *Z2z* includes a compiler that checks essential correctness properties and produces efficient code. It has been used to develop a number of gateways involving a range of issues common to protocols used in the home.

However, traditional IP-protocols are not always available for controlling networked devices. In some cases, such as X10 terminals, it is necessary to use a software library for accessing a remote device. Indeed, the sequence of library function calls to perform to complete an operation on the targeted device defines an interaction protocol. However, the protocol is often informally specified in a natural language in the library documentation.

Objet du stage

The aim of this project is twofold. First, we will define a specification language to formally define the interaction protocol supported by a library. Second, we will propose an extension of *z2z* enabling the automatic generation of gateways between a network protocol and library.

Références bibliographiques

- David Bromberg, Laurent Réveillère, Julia L. Lawall, Gilles Muller : Automatic Generation of Network Protocol Gateways. *Middleware* 2009 : 21-41