INRIA

Candidat ***Applicant***
Nom *Last Name*
MARCHAL
Prénom *First Name*
Loris

DOSSIER DE CANDIDATURE AU CONCOURS EXTERNE DE CHARGÉS DE RECHERCHE DE DEUXIÈME CLASSE POUR L'ANNÉE 2007

APPLICATION PACKET FOR THE COMPETITIVE SELECTION OF JUNIOR RESEARCH SCIENTISTS FOR YEAR 2007

DÉPÔT DES CANDIDATURES
SUBMITTING APPLICATIONS

Le dossier de candidature doit comprendre :
- Formulaire 1 : Fiche individuelle de renseignements
- Formulaire 2 : Synthèse de la candidature, en 3 pages maximum
- Formulaire 3 : Programme de recherche détaillé, en 5 pages maximum
- Formulaire 4 : Liste complète des publications
- Formulaire 5 : Lettres de recommandation
- Formulaire 6 : Déclaration de candidature

- Les lettres de recommandation, 5 au maximum
- Les rapports de thèse ou de doctorat (si disponibles)
- Une copie des derniers titres et diplômes

The application should include :
- *Form 1 : Personal information*
- *Form 2 : Application summary, maximum 3 pages*
- *Form 3 : Detailed research program, maximum 5 pages*
- *Form 4 : Complete list of publications*
- *Form 5 : Recommendation letters*
- *Form 6 : Statement of intent to apply*

- *Recommendation letters, maximum 5*
- *Ph D. dissertation reports (if available)*
- *A copy of most recent titles and diplomas*

La date limite de dépôt des dossiers de candidature est fixée au **15 février 2007**.
Les candidats doivent remettre leur **dossier en 1 exemplaire** (revêtu de la signature originale) à l'une ou plusieurs des adresses énumérées ci-dessous selon le(s) souhait(s) d'affectation :
– soit en déposant ce dossier à l'une ou plusieurs de ces adresses avant le **15 février 2007**, 16 heures ;
– soit en l'envoyant à l'une ou plusieurs de ces adresses avant le **15 février 2007** minuit, le cachet de la poste faisant foi.

The deadline to file an application *is* **February 15th, 2007**.
Applicants must supply **1 copy of their application** *(with the original signature), to one or several of the following addresses according to the research center(s) the applicant whishes to be assigned to :*
– *either by depositing this application in person at one or several of these addresses before 4 :00 PM,* ***February 15th, 2007****;*
– *or by sending this application by mail, postmarked by midnight* ***February 15th, 2007****, to one or several of these addresses.*

- Service des ressources humaines de l'unité de recherche INRIA Futurs Bordeaux,
  LABRI - Domaine Universitaire - 351 cours de la Libération, 33405 TALENCE Cedex FRANCE.
  (Téléphone/Phone : +33 (0) 5 40 00 36 39).
- Service des ressources humaines de l'unité de recherche INRIA Futurs Lille,
  LIFL - Bât.M 3 - Cité scientifique, 59655 VILLENEUVE D'ASQ cedex FRANCE.
  (Téléphone/Phone : +33 (0) 3 28 77 85 16).
- Service des ressources humaines de l'unité de recherche INRIA Futurs Saclay,
  Parc Club Orsay Université - ZAC des vignes, 4 rue Jaques Monod - Bât G, 91893 ORSAY Cedex FRANCE.
  (Téléphone/Phone : +33 (0) 1 72 92 59 23/26).
- Service des ressources humaines de l'unité de recherche INRIA Lorraine,
  Technopôle de Nancy Brabois, 615 rue du Jardin Botanique, B.P. 101, 54602 VILLERS-LES-NANCY Cedex FRANCE.
  (Téléphone/ Phone : +33 (0) 3 83 59 30 62).
- Service des ressources humaines de l'unité de recherche INRIA Rennes,
  Campus universitaire de Beaulieu, 35042 RENNES Cedex FRANCE.
  (Téléphone/Phone : +33 (0) 2 99 84 73 51/75 88).
- Service des ressources humaines de l'unité de recherche INRIA Rhône-Alpes,
  Inovallée, 655 avenue de l'Europe, Montbonnot, 38334 SAINT ISMIER Cedex FRANCE.
  (Téléphone/Phone : +33 (0) 4 76 61 54 92).
- Service des ressources humaines de l'unité de recherche INRIA Rocquencourt,
  Domaine de Voluceau, B.P. 105, 78153 LE CHESNAY Cedex FRANCE.
  (Téléphone/Phone : +33 (0) 1 39 63 57 24).
- Service des ressources humaines de l'unité de recherche INRIA Sophia-Antipolis,
  2004 Route des Lucioles, B.P. 93, 06902 SOPHIA ANTIPOLIS Cedex FRANCE.
  (Téléphone/Phone : +33 (0) 4 92 38 77 02).

Attention/ *Warning* :

Dans l'état actuel de la réglementation française, **seul le dossier original signé constitue le document officiel de candidature**[1].
According to present French regulations, ***the original application with the applicant's signature is considered as the sole official application document***[2].

Transmission du dossier de candidature par courrier électronique/ *Transmitting the application packet via e-mail*

Il est demandé au candidat d'**envoyer** le dossier de candidature[3] **par courrier électronique** (formulaires 1 à 6 dans l'ordre), **en un seul fichier**. Ce fichier, en format PDF (de préférence) ou PS sera enregistré sous le nom du candidat (`nom.prenom` ; exemple : `dupond.jean`).

Applicants are asked to ***send an electronic version***[4] *of the application packet, (with forms 1 to 6 in this order),* ***in a single file****. This file, in PDF (preferably) or PS, format is sent under the name of the applicant (*`lastname.firstname` *; for example* `dupond.jean`*).*

Ce document doit être envoyé à l'une ou plusieurs des adresses suivantes selon les souhaits d'affectation :
This document should be sent to one or several of the following addresses according to the research center(s) the applicant wishes to be assigned to :

[1] Les informations fournies par le candidat feront l'objet d'un traitement informatisé, et les listes nominatives des candidats admis à concourir, preselectionnes,admissibles et admis au concours seront accessibles sur le serveur web de l'INRIA. Le droit d'accès prévu par l'article 34 de la loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés (communication et rectification des données concernant les candidats) s'exerce auprès de la Direction des ressources humaines de l'INRIA.

[2] *The data provided in your application will be data processed. The name lists of the selected applicants will be posted on the INRIA web site. The access right as stated in art. 34 of the law N° 78.17, January 6th 1978, modified, related to data processing, files and liberty (communication and correction of the data related to your application) is filed to INRIA's Human Resources Department.*

[3] Ce document transmis par courrier électronique sera utilisé pour faciliter le travail des jurys du concours.

[4] *This document sent by e-mail will be used by the juries involved in the competitive selection process.*

cr2-bordeaux@inria.fr	pour une affectation à l'unité de recherche **Futurs Bordeaux** *for an assignment in the **Futurs Bordeaux** research center*
cr2-lille@inria.fr	pour une affectation à l'unité de recherche **Futurs Lille** *for an assignment in the **Futurs Lille** research center*
cr2-saclay@inria.fr	pour une affectation à l'unité de recherche **Futurs Saclay** *for an assignment in the **Futurs Saclay** research center*
cr2-lorraine@inria.fr	pour une affectation à l'unité de recherche **Lorraine** *for an assignment in the **Lorraine** research center*
cr2-rennes@inria.fr	pour une affectation à l'unité de recherche **Rennes** *for an assignment in the **Rennes** research center*
cr2-ralpes@inria.fr	pour une affectation à l'unité de recherche **Rhône-Alpes** *for an assignment in the **Rhône-Alpes** research center*
cr2-rocq@inria.fr	pour une affectation à l'unité de recherche **Rocquencourt** *for an assignment in the **Rocquencourt** research center*
cr2-sophia@inria.fr	pour une affectation à l'unité de recherche **Sophia-Antipolis** *for an assignment in the **Sophia-Antipolis** research center*

Formulaire 1/*Form 1*

FICHE INDIVIDUELLE DE RENSEIGNEMENTS
PERSONAL INFORMATION

Nom/*Last Name* : MARCHAL Prénom/*First Name* : Loris
Date et lieu de naissance/*Date and place of birth* : 22/02/1980, Lyon 6eme (France)
Nationalité/*Citizenship* : Français Sexe/*Sex* : M
Adresse postale/*Mailing address* : LIP - ENS Lyon
46, allée d'Italie
69364 Lyon CEDEX 07
N° de téléphone/*Telephone* : (+33) 06 64 51 58 95
Adresse électronique/*E-mail* : loris.marchal@ens-lyon.fr
Page Web personnelle/*Web page* : http://graal.ens-lyon.fr/~lmarchal/

DIPLÔMES FRANÇAIS OU ÉTRANGERS /*DIPLOMAS*

Doctorat(s)/*Ph.D.(s)* :
- Doctorat, spécialité Informatique, obtenu le 17 octobre 2006 à l'École Normale Supérieure de Lyon. Thèse intitulée « Communications collectives et ordonnancement en régime permanent sur plates-formes hétérogènes » effectuée dans l'équipe GRAAL du Laboratoire de l'Informatique du Parallélisme, UMR CNRS–ENS Lyon–UCB Lyon–INRIA 5668.

Autres diplômes (à partir du niveau maîtrise)/***Other diplomas*** *(Master's and higher)* :
- Maîtrise d'Informatique à l'École Normale Supérieure de Lyon, mention Bien, obtenue en juillet 2002.
- DEA d'Informatique Fondamentale à l'École Normale Supérieure de Lyon, mention Bien, obtenu en juillet 2003.
- Diplôme Magistère d'Informatique et Modélisation de l'École Normale Supérieure de Lyon, mention Bien, obtenu en juillet 2003.

SITUATION PROFESSIONNELLE ACTUELLE/*CURRENT PROFESSIONAL STATUS*

Statut et fonction/*Position and statute* : Allocataire de recherche et moniteur
Etablissement (ville - pays)/*Institution (city -country)* : École Normale Supérieure de Lyon – France
Date d'entrée en fonction/*Start* : 1er septembre 2003
[ ] Sans emploi /*Without employment*

FORMATION ET PARCOURS PROFESSIONNEL/*TRAINING AND PROFESSIONAL HISTORY*

Établissements français ou étrangers	Fonctions et statuts (salarié, boursier, etc.)	Dates		Observations
INSTITUTIONS French or foreign	*POSITIONS AND STATUS (employee, fellow, etc.)*	d'entrée en fonction *Start*	de cessation de fonction *End*	*REMARKS*
ÉNS-Lyon	élève normalien (fonctionnaire stagiaire)	septembre 2000	septembre 2004	
LIP/ÉNS-Lyon	allocataire de recherche / moniteur	septembre 2004	septembre 2007	

Formulaire 2/*Form 2*

SYNTHÈSE DE LA CANDIDATURE
APPLICATIONS SUMMARY

Nom/*Last name*: MARCHAL Prénom/*First name*: Loris
Projets d'affectation souhaités/Assignation whishes : GRAAL, CEPAGE.

1. Résumé de l'activité de recherche/*Summary of research activities*

Durant ma thèse, je me suis principalement intéressé aux communications collectives mises en œuvres lors de l'exécution d'applications distribuées. En supposant que le volume de données concerné est important et que les communications sont pipelinées, nous nous concentrons sur l'optimisation du régime permanent. Nous avons obtenu de nombreux résultats théoriques, qui montrent qu'il existe des algorithmes polynômiaux de débit optimal, ce qui conduit à des solutions asymptotiquement optimales en terme de temps d'exécution. Pour le cas particulier de la diffusion, nous avons montré que le débit optimal pouvait être obtenu en utilisant plusieurs arbres de diffusion, et nous avons proposé des solutions heuristiques à un seul arbre de diffusion (la recherche de l'arbre optimal étant un problème NP-complet). Ces différentes stratégies ont alors été implanté dans un outil de diffusion, que nous avons testé sur Grid5000. Ceci nous a permis de vérifier expérimentalement les résultats théoriques, ainsi que de tester plusieurs modélisation du réseau.
Je me suis également intéressé à d'autres problèmes plus classiques d'ordonnancement, comme le modèle des tâches divisibles : nous avons par exemple étudier l'influence d'une mémoire bornée pour une plate-forme en étoile. Nous avons ainsi montré qu'un problème polynômial devenait NP complet avec cette contrainte, mais que des solutions heuristiques donnaient de bons résultats dès que la mémoire disponible dépassait un seuil. Nous avons également pris en compte les messages de retour (d'ordinaire ignorés), montré que cela complexifiait le problème, et proposé des solutions optimales pour des cas simples.
Durant mon séjour post-doctoral à l'Université de Floride, j'ai contribué à développer une interface cerveau-machine. Des signaux sont captés depuis le cortex-prémoteur de rats dans un laboratoire de neuroscience (Brain Institute) puis sont acheminés jusqu'à un centre de calcul (ACIS Laboratory) où leur traitement produit une commande moteur qui est renvoyée à un robot, dans le but que l'animal puisse déplacer le robot uniquement en imaginant le mouvement. Le traitement des signaux doit être effectué avec des contraintes de temps réel. J'ai notamment contribué à intégrer un méchanisme d'apprentissage en temps réel, en m'appuyant sur une abondante littérature en traitement du signal, particulièrement sur les filtres adaptatifs et le calcul des moindres carrés. Comme détaillé dans mon projet de recherche, je souhaite pouvoir poursuivre cette collaboration en vue de paralléliser le traitement de ces données et d'étudier l'ordonnancement des différentes tâches qui en résultent.
Je participe également à l'élaboration de VoroNet, un réseau pair-à-pair permettant la recherche par le contenu. Ces travaux en cours sont développés dans la partie « projet de recherche ».

2. Résumé du programme de recherche/*Summary of research program*

Mon projet s'articule autour de trois principaux axes :
- Concevoir des ordonnancements dynamiques et distribués pour les plates-formes à grande échelle, c'est-à-dire prendre en compte et modéliser l'instabilité naturelle des plates-formes distribuées, et proposer des algorithmes robustes, décentralisés et performants pour ces plates-formes.
- L'ordonnancement pour les machines virtuelles : la virtualisation permet d'isoler des ressources pour une application et est donc très prometteuse pour le partage de ressources de calcul. Nous nous proposons d'étudier les problème de placement, de redistribution et d'ordonnancement pour ces systèmes.
- L'ordonnancement dans les réseaux : les applications de calcul distribué s'exécutent maintenant sur des réseaux longue distance (comme l'Internet), il faut un traitement particulier pour les communications volumineuses qu'elles impliquent. Il s'agit là encore de chercher des stratégies décentralisées, prenant en compte les caractéristiques de ces flux.

3. Publications/*Publications*

[A] A. Legrand, L. Marchal et Y. Robert. «Optimizing the steady-state throughput of scatter and reduce operations on heterogeneous platforms». *Journal of Parallel and Distributed Computing* **65**, numéro 12 (2005), 1497–1514. Dans cet article, nous étudions l'optimisation du débit des opérations de distribution de données (scatter) et de réduction (reduce) en régime permanent ; nous présentons des algorithmes polynomiaux asymptotiquement optimaux pour ces problèmes.

[B] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Pipelining broadcasts on heterogeneous platforms». *IEEE Trans. Parallel Distributed Systems* **16**, numéro 4 (2005). Dans cet article, nous étudions l'optimisation du débit d'une opération de diffusion en régime permanent ; nous présentons des algorithmes polynomiaux asymptotiquement optimaux pour la diffusion sur des plates-formes en arbres, en graphes acycliques dirigées ou quelconques.

[C] O. Beaumont, A.-M. Kermarrec, L. Marchal et Étienne Rivière. «VoroNet : A scalable object network based on Voronoi tessellations». Dans *International Parallel and Distributed Processing Symposium IPDPS'2007* (2007, à paraitre), IEEE Computer Society Press. Cet article présente le protocole pair-à-pair VoroNet, qui s'appuie sur un diagramme de Voronoi des objets dans l'espace de leurs caractéristiques, et permet une recherche par le contenu.

[D] H. Casanova, A. Legrand et L. Marchal. «Scheduling Distributed Applications : the SimGrid Simulation Framework». Dans *Proceedings of the third IEEE International Symposium on Cluster Computing and the Grid (CCGrid'03)* (may 2003). Cet article présente le simulateur SimGrid dans lequel a été intégré la modélisation du réseau que nous avons proposée avec Henri Casanova.

[E] L. Marchal, V. Rehn, Y. Robert et F. Vivien. «Scheduling algorithms for data redistribution and load-balancing on master-slave platforms». *Parallel Processing Letters* (2007, à paraître).

Ces publications sont disponibles sur ma page web : `http ://graal.ens-lyon.fr/~lmarchal/`

4. Réalisation et diffusion de logiciels/*Software writing and distribution*

SimGrid J'ai participé à l'élaboration et au développement dans le simulateur d'application distribuée SimGrid d'une modélisation du réseau réaliste (voir [23]). SimGrid a été créé par Henri Casanova et Arnaud Legrand, représente 15 000 lignes de code C, et est utilisé par plus d'une quarantaine de chercheurs dans le monde, ainsi que dans des buts d'éducation.

Diffusion pipelinée Pour valider les résultats théoriques obtenus sur la diffusion pipelinée en régime permanent, j'ai implanté plusieurs algorithmes et heuristiques dans un outil de diffusion. Celui-ci permet d'effectuer une diffusion sur un ensemble de machines, étant donné un graphe de topologie, en utilisant un ou plusieurs arbres de diffusion. Cet outil était à l'origine destiné à l'expérimentation, mais devant les résultats obtenus et l'intérêt de disposer d'un outil de diffusion de fichiers de grande taille, en particulier sur Grid5000, une version utilisable est en cours d'élaboration. Ce programme représente environ 5000 lignes de code C.

Interface Cerveau Machine Durant mon post-doc, j'ai contribué à développer une interface cerveau-machine. Des signaux sont captés depuis le cortex-pré-moteur de rats dans un laboratoire de neurosciences (Brain Institute) puis sont acheminés jusqu'à un centre de calcul (ACIS Laboratory) où leur traitement produit une commande moteur qui est renvoyée à un robot, dans le but que l'animal puisse déplacer le robot uniquement en imaginant le mouvement. Cette interface est encore en développement ; j'ai notamment contribué à intégrer un filtre adaptatif (utilisant l'algorithme RLS), ainsi qu'à concevoir l'architecture distribuée nécessaire pour le traitement (soit la presque totalité de l'interface côté ACIS). Le code produit représente plus de 2000 lignes en language C.

5. Valorisation et transfert technologique/*Development and technology transfer*

6. Encadrement d'activités de recherche/*Supervision of research activities*

Durant ma thèse, j'ai participé à l'encadrement de Véronika Rehn, lors de deux stages :
– stage de master 1 en 2005 : deux mois, 50% d'encadrement, co-encadré avec Yves Robert
– stage de master 2 (DEA) en 2006 : six mois, 33% d'encadrement, co-encadré avec Frédéric Vivien et Yves Robert.

Nos travaux ont portés sur l'ordonnancement de tâches divisibles avec messages de retour, puis sur l'équilibrage de charge dans les plates-formes maître-esclaves et ont été publiés dans la revue PPL [1] ainsi que les conférences PDP et HCW [10, 13].

7. Enseignement/*Teaching*

Depuis septembre 2004, je suis moniteur (titulaire d'une « allocation couplée »). J'ai donc effectué annuellement pendant ces trois dernière années 64 heures d'enseignement (équivalent TD). Ces enseignements ont été effectué en partie en classes préparatoires de l'Institut National des Sciences Appliquées (INSA) de Lyon et en partie au département d'Informatique de l'École Normale Supérieure (ENS) de Lyon. Les cours concernés étaient « Algorithmique et architectures parallèles », « Bases de données et Algorithmique », « Algorithmique des Réseaux et des Télécoms » et « Architecture, Réseaux et Système ».

8. Diffusion de l'information scientifique/*Dissemination of scientific knowledge*

9. Mobilité/*Visits*

• J'ai effectué un stage de deux mois, avant de commencer ma thèse, à l'Université de Californie, San Diego sous la direction d'Henri Casanova, en m'intéressant à la conception d'une modélisation réaliste du réseau dans un simulateur d'applications distribuées sur la grille nommée SimGrid. Cette collaboration s'est poursuivie dans le cadre l'équipe associée Inria I-ARTHUR. J'ai tout d'abord été amené à collaborer de nouveau avec Henri Casanova ainsi qu'avec son étudiant Yang Yang sur un projet d'ordonnancement de tâches divisibles sur les grilles de calcul. J'ai également collaboré avec deux autres chercheurs de l'université de San Diego, Larry Carter et Jeanne Ferrante, dans le cadre d'une étude des ordonnanceurs centralisés et décentralisés sur les plates-formes distribuées organisées en arbres. Ces diverses collaborations se sont traduites par les publications dans les conférences CCGrid et IPDPS [16, 23], ainsi que dans la revue IJHPCA [3, 12].
• En collaboration avec Pascale Vicat-Blanc Primet et Jingdi Zeng de l'équipe RESO du LIP, nous avons étudié le partage de bande passante entre requêtes pour le cœur des réseaux des grilles de calcul, ce qui a donné lieu à des publications dans les conférences GlobeComm et HPDC [11, 15].
• J'ai également collaboré avec Étienne Rivière et Anne-Marie Kermarrec de l'IRISA sur VoroNet, un projet de réseau pair-à-pair. Cette collaboration a conduit à des publications dans CSFE et IPDPS [9, 25].
• J'ai enfin effectué un séjour post-doctoral à l'Université de Floride, de septembre 2006 à février 2007, sous la direction de José Fortes : Il s'agit également de mobilité thématique, puisque dans le cadre du projet d'interface cerveau-machine, je me suis plus particulièrement intéressé à l'optimisation de l'apprentissage pour les filtres linéaires. Après une partie théorique concernant l'adaptation des méthodes existantes, j'ai participé au développement de l'interface cerveau-machine en y intégrant les algorithmes élaborés. Ces travaux font l'obet de la publication [8].

10. Responsabilités collectives/*Responsibilities*

J'ai été responsable du site Internet de la conférence HCW2004 (*workshop* satellite de IPDPS), qui servait à la fois de publicité et de plate-forme pour la soumission de manuscrits, leur relecture et leur sélection.
J'ai effectuées des relectures scientifiques pour différentes conférences (PAPP, IPDPS, Grid2005, ICPADS) et revues internationales (IJHPCA, ParCo, IEEE TPDS, JPDC).

11. Prix et distinctions/*Prizes and awards*

12. Autres éléments/*Miscellaneous*

Formulaire 3/*Form 3*

PROGRAMME DE RECHERCHE DÉTAILLÉ
DETAILED RESEARCH PROGRAM

Nom/*Last name*: MARCHAL Prénom/*First name*: Loris

Nouvelles stratégies d'ordonnancement pour le calcul distribué

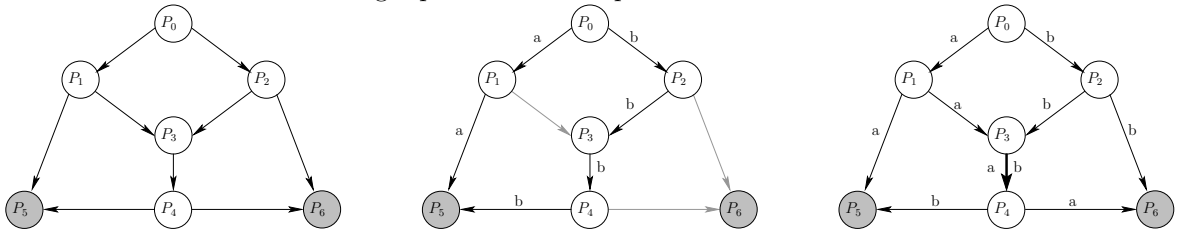
Les travaux entrepris pendant ma thèse offrent de nombreuses perspectives de recherches pour compléter et élargir les résultats obtenus. Ils soulèvent également des problématiques de recherches plus vastes que j'aimerais étudier à plus long terme. Mon projet de recherche est donc partagé entre les perspectives à court terme et à moyen terme.

Court Terme

Parmi les extensions que j'aimerais développer dans le court terme, on peut distinguer deux axes principaux.

Diffusion restreinte (multicast)

Malgré la relaxation en régime permanent, le calcul du débit d'une diffusion restreinte reste un problème NP-complet, et c'est une des rares opérations dans ce cas. On peut donc penser que la relaxation appliquée n'est pas suffisante. En particulier, si on autorise les combinaisons entre blocs diffusés, le problème semble plus simple. Considérons par exemple le réseau illustré sur la figure de gauche ci-dessous, où chaque lien peut transporter un message par unité de temps, sans contention au niveau des nœuds. La source P_0 peut envoyer un débit de deux messages simultanément à la cible P_5 : a et b représentent ces deux messages sur la figure centrale. Il en est de même pour la cible P_6, puisque le réseau est symétrique. L'approche par programmation linéaire développée dans ma thèse pourrait être étendue naïvement ici, et prédirait un débit de diffusion restreinte égal lui aussi à deux messages par unité de temps. Cependant, comme on le remarque sur la figure de droite, cela nécessiterait que deux messages soient envoyés sur le lien central (P_3, P_4), ce qui n'est pas possible dans le modèle de communication choisi. Par contre, si P_3 calcule et envoie à P_4 le ou exclusif des deux messages (a XOR b), alors P_5 et P_6 peuvent reconstruire les deux messages a et b, on obtient bien un débit de deux messages par unité de temps.



La technique illustrée sur ce petit exemple pourrait se généraliser en utilisant les résultats de la théorie du Network Coding. D'après les travaux de Ahlswede[5] et de Koetter[6], il existe un codage qui permet d'atteindre le flot maximal pour tout couple source/destination dans un graphe quelconque. Il serait intéressant d'adapter ces résultats pour essayer de montrer que la diffusion restreinte est un problème polynômial en autorisant les combinaisons (en tenant compte des coûts de calcul associés aux combinaisons). Cependant, ceci ne nous fournira pas nécessairement de méthode utilisable en pratique. On pourrait alors s'inspirer des travaux sur l'utilisation de combinaisons aléatoires[7] : ceux-ci montrent que si chaque nœud choisit de façon aléatoire les combinaisons de blocs qu'il va diffuser, alors on peut reconstruire le message total en chaque nœud cible

[5] Ahlswede et al., *Network information flow* IEEE Transactions on Information Theory, 2000.
[6] Ho et al. *An information theoretic view of network management*, INFOCOM, 2003.
[7] Gkantsidis et al. *Network coding for large scale content distribution*, In INFOCOM, 2005.

avec forte probabilité. Si cette méthode peut être utilisée ici, elle permettrait d'éviter un contrôle centralisé et coûteux des combinaisons. Si cette technique paraît indispensable pour la diffusion restreinte, elle peut également être utile pour la diffusion totale, en évitant la construction centralisée d'arbres de diffusion concurrents.
Un des problèmes qui peut limiter l'utilisation de combinaisons de blocs est la puissance de calcul nécessaire pour (i) créer de nouvelles combinaisons (en particulier, même les nœuds qui ne sont pas des destinations doivent créer des combinaisons, et on voudrait leur éviter une grosse charge de calcul pour une opération dont il ne font pas partie) et (ii) reconstituer le message initial à partir de combinaisons, ce qui demande une inversion de matrice de grande dimension, à coefficients dans des corps finis de grande taille. Il s'agit alors d'étudier le compromis entre le bénéfice de cette approche et la surcharge de calcul nécessaire.

Topologies pour le calcul distribué

Dans la plupart des travaux précédents, nous supposons connaître parfaitement le graphe de communication, et de pouvoir contrôler tous les nœuds de la plate-forme : même dans la diffusion restreinte (multicast), on suppose que les nœuds qui ne sont pas des cibles de la diffusion participent en routant les messages de la façon souhaitée par l'ordonnancement. Cette hypothèse est valable pour des réseaux locaux et/ou privés, mais devient de moins en moins pertinente lorsqu'on s'intéresse à des réseaux à grande échelle, surtout lorsque les liens de communication sont partagés entre de nombreux utilisateurs. Nous avons d'ailleurs proposé une modélisation simple du partage de bande-passante dans les liens longue-distance, pour ordonnancer des applications divisibles sur une plate-forme utilisant plusieurs sites de calcul reliés par de tels liens. Cette modélisation convient pour une plate-forme simple constituée de sites (grappes de calcul) reliés par des liens longue-distance, mais elle ne permet pas d'utiliser les techniques d'ordonnancement de communications collectives que nous avons développées (le problème d'ordonnancement simple étudié sur cette plate-forme est déjà NP-complet).
On pourrait également essayer de reconstruire le graphe de plate-forme que nous utilisons. Cependant, acquérir une information complète de la topologie est une tâche longue et ardue. Pour connaître précisément le graphe de la plate-forme, il est *a priori* nécessaire de vérifier pour toute paire de routes ($P_i \to P_j$ et $P_k \to P_l$) si des transferts concurrents sur ces deux routes interfèrent. De plus, il faut être capable de mesurer ou d'estimer la bande-passante de chaque lien de communication dans le graphe ainsi construit. Plutôt que d'utiliser un tel graphe « exhaustif » de la plate-forme, il serait intéressant de pouvoir obtenir une modélisation approchée mais plus « utilisable » de la plate-forme, en ne cherchant pas à modéliser précisément les parties du réseau qui nous sont inaccessibles : le réseau interne (longue-distance) est souvent sur-dimensionné ; on peut dans ce cas se contenter d'une analyse locale.
Il serait enfin très intéressant de se tourner vers d'autres topologies naturellement adaptées à des environnement à grande-échelle, comme les topologies pair-à-pair. Celles-ci ont fait leurs preuves pour le partage de données de grande taille, sur des environnements distribués à grande échelle. Un réseau pair-à-pair serait particulièrement adapté pour gérer des environnements de calcul participatif (comme les projets seti@home, « Berkeley Open Infrastructure for Network Computing » ou « World Community Grid »). En général, les réseaux pair-à-pair ont de bonnes propriétés de tolérance aux pannes, de stabilité, et de passage à l'échelle, grâce à l'utilisation d'un réseau virtuel (*overlay*) dont la topologie est bien connue. Les opérations possibles sur ces réseaux se limitent souvent à la recherche ou la diffusion de données. De même la description des pairs est très simple : ils sont le plus souvent tous équivalents, et quelque fois munis de bande-passante d'entrée et de sortie. Il faudrait donc adapter ces topologies virtuelles afin de pouvoir concevoir des ordonnancements sur ces plates-formes, en particulier pour l'exécution de tâches indépendantes, qui constituent une application naturelle pour ce type de plates-formes distribuées à grande échelle.

Moyen Terme

À plus long terme, mon projet de recherche s'articule autour de trois thèmes, chacun en coopération soit avec une équipe du Laboratoire de l'Informatique du Parallélisme (ENS Lyon), soit avec un laboratoire extérieur.

Gestion de la dynamicité et algorithmes décentralisés

Nous avons jusqu'à présent étudié les plates-formes hétérogènes et conçu des algorithmes statiques pour ces plates-formes. Nous savons relativement bien modéliser et utiliser de façon performante ces plates-formes à l'aide de divers outils théoriques, comme la modélisation des applications en tâches indépendantes, en tâches divisibles ou encore l'ordonnancement en régime permanent.
Cependant ces algorithmes sont valides sur des plates-formes dont les caractéristiques ne changent pas au cours du temps. Or les plates-formes réelles que nous souhaitons utiliser ne vérifient pas cette hypothèse : si les plates-formes de taille moyenne (de type grappes de calcul) sont à peu près statiques, les plates-formes à grande échelle (grappes de grappes, grilles de calcul) sont fondamentalement dynamiques. Les performances des liens de communication ne sont pas constantes car il existe sur ces liens un trafic extérieur que l'on ne maîtrise pas et qui est difficile à modéliser. La vitesse des unités de calcul est également dynamique car affectée par une charge extérieure. Sur de telles plates-formes, il est également réaliste de considérer que certaines unités de calcul vont subir de défaillances pendant leur utilisation. Il nous faut donc concevoir des algorithmes robustes aux fluctuations et tolérants aux pannes, qui puissent s'adapter à des ressources dynamiques, voire volatiles.
Avant de concevoir des algorithmes pour les plates-formes dynamiques, un premier travail de modélisation est nécessaire. Il faut modéliser la variation de charge, le comportement des utilisateurs sur des machines partagées, le trafic du réseau, etc. La plupart des outils disponibles pour modéliser des ressources dynamiques ne sont pas adaptés à notre approche. D'autre part, les lois de distribution de probabilités généralement utilisées sont des lois de Poisson ou plus souvent de distribution exponentielle qui ont l'avantage d'être manipulables. Cependant, les récents travaux de Feitelson[8] sur la modélisation de la charge des plates-formes et du comportement des utilisateurs montrent que d'autres distributions sont plus appropriées.
Outre la modélisation de l'instabilité de la plate-forme, il nous faut également mettre au point une métrique adaptée pour comparer les performances de différents ordonnancements. On peut imaginer qu'un algorithme robuste est un algorithme performant en moyenne, étant données les distributions de probabilités des performances des ressources ; au contraire, ce peut être un algorithme garanti sur tout un domaine de conditions possibles.
De là, nous pouvons imaginer concevoir deux types d'algorithmes. On peut choisir d'élaborer des algorithmes résistants, qui ne sont pas sensibles aux variations de charge et les absorbent naturellement. Dans ce cadre là, un algorithme statique (ne modifiant pas ses décisions d'ordonnancement au cours du temps) peut être adapté à une plate-forme dynamique, probablement si les variations de performances sont limitées. Pour des variations plus importantes ou pour faire face à des pannes, il est nécessaire de concevoir des algorithmes réactifs, qui construisent l'ordonnancement à la volée ou remettent en cause l'ordonnancement initial.
Une approche pour aboutir à des algorithmes s'adaptant aux plates-formes dynamiques est de concevoir des algorithmes résistant localement aux variations de performances, afin d'obtenir une garantie globale. On peut par exemple penser aux algorithmes utilisant une stratégie d'équilibrage local qui permet d'obtenir un équilibre global. C'est ce que nous avons commencé à faire en utilisant l'algorithme d'Awerbuch et Leighton[9] qui, par des équilibrages locaux sur des files d'attente, offre une garantie de convergence sous des conditions dynamiques.
On peut ensuite imaginer ajouter à l'ordonnanceur distribué une propagation des informations de variations de performances, afin d'améliorer la vitesse de convergence, en s'appuyant éventuellement sur une vision hiérarchique (ou au moins organisée) de la plate-forme : une information détectée au sein d'une grappe de calcul sera propagée au reste de la plate-forme pour que les ordonnanceurs locaux s'y adaptent. Si on peut espérer concevoir un schéma expérimental de validation de tels algorithmes, il est sans doute beaucoup plus difficile de prédire et de garantir le comportement des ordonnancements, même si on peut espérer obtenir des résultats théoriques dans des cas simples.
L'intérêt des algorithmes adaptatifs décentralisés est double : ils permettent de garantir des performances acceptables malgré la dynamicité des plates-formes et ils offrent des propriétés de passage à l'échelle, qui les

[8] Dror G. Feitelson, *Workload Characterization and Modeling*, http://www.cs.huji.ac.il/~feit/wlmod/
[9] B. Awerbuch and T. Leighton, *Improved approximation algorithms for the multi-commodity flow problem and local competitive routing in dynamic networks*, IEEE Symposium on Foundations of Computer Science, 1994.

autorisent à être mis en œuvre sur de larges plates-formes.
L'étude des topologies pair-à-pair pour le calcul distribué proposée dans la partie « Court terme » trouve naturellement son prolongement dans l'étude des algorithmes décentralisés et robustes : nous espérons que les connaissances aquises sur ces topologies et les structures de données distribuées nous permettrons de développer des algorithmes dynamiques et performants.
Ce projet d'étude des algorithmes dynamiques et décentralisés s'inscrit naturellement dans le cadre de l'équipe GRAAL du LIP.

Ordonnancement de machines virtuelles

Lors de mon séjour post-doctoral au laboratoire ACIS de l'université de Floride, j'ai été amené à considérer des problèmes d'ordonnancement pour les machines virtuelles. L'application à laquelle nous nous intéressons est une interface cerveau-machine : elle reçoit périodiquement des signaux provenant du cerveau d'un animal et doit produire une commande moteur envoyée à un robot. Le but ultime est de permettre à des personnes handicapées de déplacer des membres artificiels uniquement par la pensée. Le traitement des données provenant du cerveau consiste en un mélange d'avis de « processus experts » : un grand nombre de modèles indépendants (les experts) produisent chacun une réponse. Ces résultats sont agrégés en une réponse finale en tenant compte de l'importance des différents modèles, les importances étant mises à jour périodiquement. Il existe également un processus d'apprentissage qui permet à chaque modèle de raffiner ses réponses en mettant à jour ses paramètres. Durant chaque période, les modèles désignés comme importants doivent calculer leur réponse en un temps très limité, alors que autres calculs (les autres modèles et l'apprentissage) peuvent s'exécuter plus lentement.
Pour exécuter ces calculs, nous comptons utiliser l'architecture In-Vigo[10] développée à l'université de Floride. Celle-ci permet de créer de façon dynamique un ensemble de ressources virtuelles pour une application. Les tâches correspondants aux différents « experts » peuvent ainsi être exécutées comme des processus communicants au sein d'une même machine virtuelle, ou encore comme différentes machines virtuelles. Certaines de ces tâches doivent être exécutées rapidement, le temps de traitement ne devant pas excéder quelques centaines de millisecondes, alors que d'autres tâches ont des contraintes de temps d'exécution plus souples.
Plusieurs problèmes d'ordonnancement doivent être résolus. Il faut savoir ordonnancer précisément les différentes tâches présentes sur une machine physique afin de garantir le temps d'exécution des tâches importantes. Il faut également savoir répartir les différentes tâches entre les machines disponibles, pour en même temps minimiser le nombre de machines utilisées (et le temps de communications entre elles) et garantir le temps d'exécution de certaines tâches. Pour ceci, les modifications apportées à l'ordonnanceur en charge des machines virtuelles cohabitant sur un même hôte physique doivent être prises en compte pour modéliser le comportement d'une machine. Enfin, il faut aussi savoir redistribuer ces différentes tâches : comme le système évolue, l'importance des différents experts change, certaines tâches deviennent donc plus prioritaires, d'autres moins prioritaires, et le nombre d'experts (et donc de tâches) peut aussi varier. Les algorithmes existants de rééquilibrage de charge doivent donc être adaptés afin de prendre en compte les contraintes de temps-réel spécifiques à cette application.
Si ce travail concerne principalement une application donnée de traitement du signal pour les neurosciences, l'ordonnancement pour machines virtuelles est également un sujet d'actualité pour d'autres applications. Les machines virtuelles sont en particulier utilisées pour des applications de services Web : on peut citer l'exemple d'Amazon qui met ses serveurs à disposition d'utilisateurs extérieurs pour héberger des machines virtuelles exécutant des services Web dans son « Amazon Elastic Compute Cloud[11] ». Là encore, il faut réfléchir aux stratégies de placement des différentes machines virtuelles sur les machines physiques, et au problème de la redistribution : la charge des différents servives hébergés évolue, et avec elle les ressources nécessaires aux différentes machines virtuelles. Un service qui voit son trafic augmenter peut être amené à se dupliquer. On peut alors imaginer concevoir une collaboration entre l'ordonnanceur du service (qui connaît la charge et peut décider de la duplication) et l'ordonnanceur des machines virtuelles qui gère l'ensemble de serveurs disponibles.

[10] S. Adabala et al, *From Virtualized Resources to Virtual Computing Grids : The In-VIGO System*, FGCS 21(6), 2005.
[11] http://aws.amazon.com/ec2

Ce travail se poursuivra naturellement en collaboration avec le laboratoire ACIS de l'université de Floride.

Ordonnancement pour les réseaux

Le troisième axe de mon projet de recherche est en collaboration avec l'équipe RESO du Laboratoire de l'Informatique du Parallélisme. Il consiste en une poursuite des travaux déjà entrepris avec cette équipe (et publiés dans les conférences GlobeCom 2005 et HPDC 2006).
Nous partons de la constatation qu'actuellement, dans l'utilisation des réseaux à grande échelle, aucune anticipation n'est faite pour ordonnancer et gérer les différents flux se partageant les ressources disponibles. Des informations sur la charge et la nature du trafic sont bien utilisées par les opérateurs pour « provisionner » le réseau, c'est-à-dire pour connaître le nombre et la localisation des équipements nécessaires lors de la construction ou de l'évolution d'un réseau, mais aucune information explicite de charge n'est utilisée lors du routage et de l'acheminement des différents flux.
Or aujourd'hui, il devient possible de prédire l'utilisation du réseau par certaines applications, en particulier pour des applications de calcul scientifique distribuées. Lors du lancement d'une telle application, on peut connaître assez précisément ses besoins en communications. Ces besoins sont d'une nature différente des autres flux utilisant le réseaux : il s'agit généralement de gros volumes de données à transférer sans taux de transmission imposé, mais qui doivent être effectués à l'intérieur d'une fenêtre temporelle prédéfinie (afin de respecter les dépendances de données et les contraintes de synchronisation de l'application).
En collaboration avec l'équipe RESO, nous avons déjà commencé à étudier ce problème pour le cas de topologies simples, en supposant que le cœur du réseau était sur-dimensionné et que la seule contention possible avait lieu au niveau des points d'accès (entrants ou sortants). Nous souhaiterions maintenant nous intéresser à des topologies de réseaux plus générales. Un grand nombre d'applications de calcul scientifique ne sont en effet plus destinées à être exécutées sur des environnements dédiés, mais doivent s'adapter à des plates-formes ou la ressource réseau est partagée et peut constituer un goulet d'étranglement.
De nombreux travaux existent dans le domaine des réseaux qui visent à la réservation de ressources et/ou à la qualité de service. On peut par exemple citer le protocole de réservation RSVP[12] qui, immédiatement avant l'exécution d'un transfert, réserve la bande passante souhaitée le long de la route empruntée. Cependant, aucune information sur les besoins des applications n'est signalée à l'avance, alors que cette information est disponible : nous voulons donc la prendre en compte afin d'optimiser l'utilisation des ressources. D'autres approches comme RON (*Resilient Overlay Network*[13]) créent un réseau logique (*overlay*) par dessus le réseau physique (Internet) pour assurer que les routes reliant les sites utilisés ont les propriétés requises. L'inconvénient d'une telle approche est que pour assurer un « routage utilisateur », les transferts doivent « remonter » vers chacun des sites présents sur leur route. Or la congestion se trouve le plus souvent au niveau de la connexion des sites au réseau longue-distance, qui est une partie critique pour les performances de cette approche. On peut également citer d'autres travaux plus spécifiques comme la réservation de longueur d'ondes dans les réseaux optiques pour des plates-formes dédiées, ou encore l'utilisation de plusieurs supports de communication (optique, IP, sans-fil, etc.) en fonction du type de trafic en vue de garantir une qualité de service (QoS-routing[14]).
En vue de réserver des ressources pour des transferts anticipés, nous pouvons donc agir sur la dimension spatiale : en réservant des ressources différentes, comme des routes avec une qualité de service garantie, pour un trafic spécial. En anticipant les requêtes, nous pouvons également agir au niveau temporel, en modifiant le taux de transfert d'une application au cours du temps. Nous voulons également concevoir des algorithmes qui offrent une certaine forme de robustesse et peuvent faire face à des variations du trafic tout en étant adaptés à des plates-formes de grande taille. Il s'agit là encore d'utiliser des algorithmes décentralisés pour résoudre ces problèmes. Le protocole de calcul des plus courts chemins OSPF[15] est un bon exemple d'algorithme décentralisé dont nous voudrions nous inspirer pour concevoir des stratégies de réservation et d'ordonnancement de flux qui passe à l'échelle.

[12] R. Braden et al., *Resource ReSerVation Protocol (RSVP)*, RFC 2205, September 1997, Proposed Standard.
[13] D. G. Andersen et al., *The Case for Resilient Overlay Networks*, HotOS-VIII, May 2001
[14] P. Paul et S. V. Raghavan, *Survey of QoS routing*, ICCC 2002.
[15] John T. Moy, *Anatomy of an Internet Routing Protocol*, Addison-Wesley Professional, 1998

Formulaire 4/*Form 4*

LISTE COMPLÈTE DES PUBLICATIONS[16]
COMPLETE PUBLICATION LIST[17]

Nom/*Last name*: MARCHAL Prénom/*First name*: Loris

Revues internationales avec comité de lecture

[1] L. Marchal, V. Rehn, Y. Robert et F. Vivien. «Scheduling algorithms for data redistribution and load-balancing on master-slave platforms». *Parallel Processing Letters* (2007, à paraître).

[2] O. Beaumont, L. Marchal et Y. Robert. «Complexity results for collective communications on heterogeneous platforms». *Int. Journal of High Performance Computing Applications* **20**, numéro 1 (2006).

[3] L. Marchal, Y. Yang, H. Casanova et Y. Robert. «Steady-state scheduling of multiple divisible load applications on wide-area distributed computing platforms». *Int. Journal of High Performance Computing Applications* **20**, numéro 3 (2006).

[4] A. Legrand, L. Marchal et Y. Robert. «Optimizing the steady-state throughput of scatter and reduce operations on heterogeneous platforms». *Journal of Parallel and Distributed Computing* **65**, numéro 12 (2005), 1497–1514.

[5] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Steady-state scheduling on heterogeneous clusters». *Int. J. of Foundations of Computer Science* **16**, numéro 2 (april 2005).

[6] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Pipelining broadcasts on heterogeneous platforms». *IEEE Trans. Parallel Distributed Systems* **16**, numéro 4 (2005).

[7] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Scheduling strategies for mixed data and task parallelism on heterogeneous clusters». *Parallel Processing Letters* **13**, numéro 2 (2003).

Conférences internationales avec comité de lecture

[8] J. DiGiovanna, L. Marchal, P. Rattanatamrong, M. Zhao, S. Darmanjian, B. Mahmoudi, J. Sanchez, J. Príncipe, L. Hermer-Vazquez, R. Figueiredo et J. Fortes. «Towards Real-Time Distributed Signal Modeling for Brain Machine Interfaces». Dans *Dynamic Data Driven Application Systems, workshop satellite de ICCS 2007* (2007, à paraitre), Springer Verlag LNCS.

[9] O. Beaumont, A.-M. Kermarrec, L. Marchal et Étienne Rivière. «VoroNet : A scalable object network based on Voronoi tessellations». Dans *International Parallel and Distributed Processing Symposium IPDPS'2007* (2007, à paraitre), IEEE Computer Society Press.

[10] V. Rehn, Y. Robert et F. Vivien. «Scheduling and data redistribution strategies on star platforms». Dans *PDP'2007, 15th Euromicro Workshop on Parallel, Distributed and Network-based Processing* (2007), IEEE Computer Society Press.

[11] L. Marchal, P. V.-B. Primet, Y. Robert et J. Zeng. «Optimal Bandwidth Sharing in Grid Environment». Dans *15th International Symposium on High Performance Distributed Computing (HPDC 2006* (2006), IEEE Computer Society Press.

[12] O. Beaumont, L. Carter, J. Ferrante, A. Legrand, L. Marchal et Y. Robert. «Centralized versus distributed schedulers for multiple bag-of-task applications». Dans *International Parallel and Distributed Processing Symposium IPDPS'2006* (2006), IEEE Computer Society Press.

[16] Les publications les plus significatives devront être consultables sur la page web du candidat.
[17] *Most relevant publications have to be available for consultation via the web page of the applicant.*

[13] O. Beaumont, L. Marchal, V. Rehn et Y. Robert. «FIFO scheduling of divisible loads with return messages under the one-port model». Dans *HCW'2006, the 15th Heterogeneous Computing Workshop* (2006), IEEE Computer Society Press.

[14] O. Beaumont, L. Marchal et Y. Robert. «Scheduling divisible loads with return messages on heterogeneous master-worker platforms». Dans *International Conference on High Performance Computing HiPC'2005* (2005), volume 3769 des *LNCS*, Springer Verlag, pp. 498–507.

[15] L. Marchal, P. V.-B. Primet, Y. Robert et J. Zeng. «Optimizing Network Resource Sharing in Grids». Dans *IEEE Global Telecommunications Conference (Gloebcom'2005)* (2005, to appear), IEEE Computer Society Press.

[16] L. Marchal, Y. Yang, H. Casanova et Y. Robert. «A realistic network/application model for scheduling divisible loads on large-scale platforms». Dans *International Parallel and Distributed Processing Symposium IPDPS'2005* (2005), IEEE Computer Society Press.

[17] O. Beaumont, L. Marchal et Y. Robert. «Broadcast Trees for Heterogeneous Platforms». Dans *International Parallel and Distributed Processing Symposium IPDPS'2005* (2005), IEEE Computer Society Press.

[18] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Independent and Divisible Tasks Scheduling on Heterogeneous Star-schaped Platforms with Limited Memory». Dans *13th Euromicro Conference on Parallel, Distributed and Network-based Processing PDP'2005* (2005), IEEE Computer Society Press, pp. 179–186.

[19] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Pipelining broadcasts on heterogeneous platforms». Dans *International Parallel and Distributed Processing Symposium IPDPS'2004* (2004), IEEE Computer Society Press.

[20] A. Legrand, L. Marchal et Y. Robert. «Optimizing the steady-state throughput of scatter and reduce operations on heterogeneous platforms». Dans *APDCM'2004, 6th Workshop on Advances in Parallel and Distributed Computational Models* (2004), IEEE Computer Society Press.

[21] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Complexity results and heuristics for pipelined multicast operations on heterogeneous platforms». Dans *Proceedings of the 33rd International Conference on Parallel Processing (ICPP'04)* (2004), IEEE Computer Society Press.

[22] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Steady-state scheduling on heterogeneous clusters : why and how ?». Dans *6th Workshop on Advances in Parallel and Distributed Computational Models APDCM 2004* (2004), IEEE Computer Society Press.

[23] H. Casanova, A. Legrand et L. Marchal. «Scheduling Distributed Applications : the SimGrid Simulation Framework». Dans *Proceedings of the third IEEE International Symposium on Cluster Computing and the Grid (CCGrid'03)* (may 2003).

[24] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Assessing the impact and limits of steady-state scheduling for mixed task and data parallelism on heterogeneous platforms». Dans *HeteroPar'2004 : International Conference on Heterogeneous Computing, jointly published with ISPDC'2004 : International Symposium on Parallel and Distributed Computing* (2004), IEEE Computer Society Press.

Conférences nationales avec comité de lecture

[25] O. Beaumont, A.-M. Kermarrec, L. Marchal et Étienne Rivière. «Voronet, un réseau objet-à-objet sur le modèle petit-monde». Dans *CFSE'5 : Conférence Française sur les Systèmes d'Exploitation* (2006).

Rapports de recherche

[26] L. Marchal, V. Rehn et F. Vivien. «Scheduling and data redistribution strategies on star platforms». Research report, LIP, ENS Lyon, France, juin 2006.

[27] O. Beaumont, A.-M. Kermarrec, L. Marchal et E. Rivière. «VoroNet : A scalable object network based on Voronoi tessellations». Research report, LIP, ENS Lyon, France, février 2006.

[28] O. Beaumont, L. Marchal, V. Rehn et Y. Robert. «FIFO scheduling of divisible loads with return messages under the one-port model». Research report, LIP, ENS Lyon, France, octobre 2005.

[29] O. Beaumont, L. Carter, J. Ferrante, A. Legrand, L. Marchal et Y. Robert. «Scheduling multiple bags of tasks on heterogeneous master-worker platforms : centralized versus distributed solutions». Research report, LIP, ENS Lyon, France, septembre 2005.

[30] L. Marchal, P. V.-B. Primet, Y. Robert et J. Zeng. «Scheduling network requests with transmission window». Research report, LIP, ENS Lyon, France, juillet 2005.

[31] O. Beaumont, L. Marchal et Y. Robert. «Scheduling divisible loads with return messages on heterogeneous master-worker platforms». Research report, LIP, ENS Lyon, France, mai 2005.

[32] L. Marchal, P. V.-B. Primet, Y. Robert et J. Zeng. «Optimizing Network Resource Sharing in Grids». Research report, LIP, ENS Lyon, France, mars 2005. Also available as INRIA Research Report RR-5523.

[33] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Complexity results and heuristics for pipelined multicast operations on heterogeneous platforms». Research report, LIP, ENS Lyon, France, février 2004. Also available as INRIA Research Report RR-5123.

[34] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Steady-State Scheduling on Heterogeneous Clusters : Why and How ?». Research report, LIP, ENS Lyon, France, mars 2004.

[35] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Assessing the impact and limits of steady-state scheduling for mixed task and data parallelism on heterogeneous platforms». Research report, LIP, ENS Lyon, France, avril 2004. Also available as INRIA Research Report RR-5198.

[36] L. Marchal, Y. Yang, H. Casanova et Y. Robert. «A realistic network/application model for scheduling divisible loads on large-scale platforms». Research report, LIP, ENS Lyon, France, avril 2004. Also available as INRIA Research Report RR-5197.

[37] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Independent and Divisible Task Scheduling on Heterogeneous Star-shaped Platforms with Limited Memory». Research report, LIP, ENS Lyon, France, avril 2004. Also available as INRIA Research Report RR-5196.

[38] O. Beaumont et L. Marchal. «Pipelining broadcasts on heterogeneous platforms under the one-port model». Research Report RR-2004-32, LIP, ENS Lyon, France, juillet 2004.

[39] O. Beaumont, L. Marchal et Y. Robert. «Broadcast Trees for Heterogeneous Platforms». Research Report RR-2004-46, LIP, ENS Lyon, France, novembre 2004.

[40] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Optimal algorithms for the pipelined scheduling of task graphs on heterogeneous systems». Research Report RR-2003-29, LIP, ENS Lyon, France, avril 2003. Also available as INRIA Research Report RR-4870.

[41] A. Legrand, L. Marchal et Y. Robert. «Optimizing the steady-state throughput of scatter and reduce operations on heterogeneous platforms». Research Report RR-2003-33, LIP, ENS Lyon, France, juin 2003. Also available as INRIA Research Report RR-4872.

[42] O. Beaumont, A. Legrand, L. Marchal et Y. Robert. «Optimizing the steady-state throughput of broadcasts on heterogeneous platforms heterogeneous platforms». Research report, LIP, ENS Lyon, France, juin 2003. Also available as INRIA Research Report RR-4871.

[43] H. Casanova et L. Marchal. «A Network Model for Simulation of Grid Application». Research Report RR-2002-40, LIP, ENS Lyon, France, octobre 2002. Also available as INRIA Research Report RR-4596.

Formulaire 5/*Form 5*

LETTRES DE RECOMMANDATION
RECOMMENDATION LETTERS
5 NOMS AU MAXIMUM/*MAXIMUM 5 NAMES*

Nom du candidat/*Applicant's Last Name*: MARCHAL Prénom/*First name*: Loris

Le candidat est invité à joindre au dossier de candidature les originaux des lettres de recommandation qui lui auront été adressées par des personnalités du milieu académique ou industriel./The candidat may enclose the original of recommendation letters written by references from academia or industry.

Noms et adresses (inclure l'adresse électronique)/*Names and addresses (including email)* :

1. **Prof. Yves ROBERT**
   École Normale Supérieure de Lyon
   LIP - ÉNS Lyon
   46, allée d'Italie
   69364 Lyon CEDEX 07, FRANCE
   Phone : (+33) (0)4 72 72 85 86
   Email : `yves.robert@ens-lyon.fr`
2. **Prof. Pascale VICAT-BLANC PRIMET**
   Directrice de recherche INRIA
   École Normale Supérieure de Lyon
   LIP - ÉNS Lyon
   46, allée d'Italie
   69364 Lyon CEDEX 07, FRANCE
   Phone : (+33) (0)4 72 72 88 02
   Email : `pascale.primet@inria.fr`
3. **Prof. Henri CASANOVA**
   University of Hawaii at Manoa
   Pacific Ocean Science and Technology Building, Room #317
   1680 East-West Road, Honolulu, HI 96822, USA
   Phone : +1 (808) 956 2649
   Email : `henric@hawaii.edu`
4. **Prof. Jeanne FERRANTE**
   Professor and Associate Dean of the Jacobs School of Engineering
   University of California at San Diego
   CSE Building, Room 3102
   9500 Gilman Drive,
   La Jolla, CA 92093-0404, USA
   Phone : +1 (858) 534 8406
   Email : `ferrante@cs.ucsd.edu`
5. **Dr. José A. B. FORTES**
   Professor and BellSouth Eminent Scholar
   Director, Advanced Computing and Information Systems (ACIS) Laboratory
   University of Florida
   Department of Electrical and Computer Engineering
   P.O. Box 116200, 339 Larsen Hall, Gainesville, FL 32611-6200, USA
   Phone : +1 (352) 392 9265
   Email : `fortes@acis.ufl.edu`

Formulaire 6/*Form 6*

DÉCLARATION DE CANDIDATURE
STATEMENT OF INTENT TO APPLY

Je soussigné(e)/*I, the undersigned* MARCHAL Loris déclare présenter ma candidature au concours de recrutement de chargés de recherche de deuxième classe de l'INRIA pour l'année 2007/*hereby declare that I apply for the 2007 competitive selection for INRIA junior research scientists (chargés de recherche de deuxième classe) positions.*

Mon programme de recherche s'intitule/*Title of my research program*

Nouvelles stratégies d'ordonnancement pour le calcul distribué

En cas de réussite au concours je demande à être affecté(e) au sein du (ou des) projets de recherche suivants[18]/*If I am recruited by INRIA I wish to be assigned to the following research project-team(s)*[19] :

Les candidats sont invités à prendre contact avec les chefs des projets dans lesquels ils postulent/Applicants should enter in contact with the project leaders concerned by their applications.

	Projet de recherche *Project-team*
[X] concours FUTURS BORDEAUX	CEPAGE
[ ] concours FUTURS LILLE	
[ ] concours FUTURS SACLAY	
[ ] concours LORRAINE	
[ ] concours RENNES	
[X] concours RHÔNE-ALPES	GRAAL
[ ] concours ROCQUENCOURT	
[ ] concours SOPHIA ANTIPOLIS	

J'ai pris connaissance des conditions requises pour concourir[20], et je certifie sur l'honneur l'exactitude des renseignements fournis dans ce dossier/*I am aware of the conditions*[21] *required for the consideration of my application and I certify that the information I have supplied is true and correct.*

[18] Inscrire une croix dans la ou les cases choisies. Les chargés de recherche de deuxième classe de l'INRIA sont recrutés au sein de l'un des projets de recherche existants (ou en cours de création au moment du concours). C'est pourquoi il est demandé aux candidats d'indiquer le ou les projets de recherche auxquels ils souhaitent être rattachés en cas de recrutement (le nombre de projets de recherche indiqués ne doit pas excéder 2). Pour chaque projet de recherche mentionné, indiquer l'unité de recherche considérée : Futurs Bordeaux,Futurs Lille, Futurs Saclay, Lorraine, Rennes, Rocquencourt, Rhône-Alpes ou Sophia-Antipolis ; si le candidat postule à un projet localisé dans deux unités de recherche, il doit mentionner la ou les unités de recherche choisies. Voir la liste des projets de l'INRIA sur `http ://www.inria.fr/recherche/equipes/listes/index.fr.html`. Dans le cadre des souhaits émis par le candidat, la direction se réserve le droit de choisir l'unité de recherche d'accueil.

[19] *Check one or more boxes. INRIA junior research scientists (chargés de recherche de deuxième classe) are recruited within one of the existing project-teams (or in one of the project-teams being currently under creation). The applicant is asked to indicate the project-team(s) he or she wishes to be assigned to (no more than 2 project-teams). For each research project-team mentioned, indicate the research center : Futurs Bordeaux, Futurs Lille, Futurs Saclay, Lorraine, Rennes, Rocquencourt, Rhônes-Alpes or Sophia-Antipolis. If the applicant is applying to a project-team based in two research centers, the chosen research center(s) must be mentioned. See the list of INRIA research project-teams on* `http ://www.inria.fr/recherche/equipes/listes/index.en.html`. *As part of the wishes expressed by the candidate, the management reserves the right to choose the research center assigned.*

[20] Voir la brochure d'information.

[21] *See the information booklet.*

À/*City* Lyon, le/*Date* 15/02/2007
Signature