

Rapport sur le manuscrit proposé par
Stéphane Génaud en vue de l'obtention de
l'habilitation à diriger des recherches de
l'université Henri Poincaré – Nancy 1

Christophe Cérin

Vendredi 30 octobre 2009

Préambule

Le manuscrit proposé est constitué de trois parties principales et d'une introduction sur deux pages qui présente le document. Nous allons partir de la fin du document, c'est à dire des annexes qui contiennent des copies d'articles les plus significatifs et du curriculum vitæ.

Analyse du Curriculum Vitæ

Le candidat est âgé de 40 ans et il est sur le domaine du parallélisme depuis 1992-1993 où il a obtenu un DESS dans cette discipline, puis une thèse à Strasbourg en 1997. Il a été nommé Maître de conférences dans la foulée (1998) et à Strasbourg. Entre 2007 et 2009 il a été détaché en qualité de Chargé de recherche au Loria de Nancy, dans l'équipe projet INRIA AlGorille. Depuis septembre 2009 il est de retour à Strasbourg et en qualité de Maître de conférences. Il a donc été mobile au cours de ces années.

La lecture du document fait apparaître un effort du candidat pour valoriser son travail dans de bonnes revues internationales (Journal of Grid Computing, Journal of SuperComputing, Parallel Computing) ainsi qu'un chapitre de livre. Les publications en conférences internationales sont d'un bon niveau et elles sont clairement le résultat d'une partie du travail d'encadrement de thèses qui de toute évidence a porté ses fruits. On peut compter une vingtaine de publications en anglais, en conférences internationales ou en revues internationales ce qui témoigne d'une activité soutenue. Au total son curriculum fait état de 31 publications. Il a encadré ou co-encadré 3 thèses à ce jour (à 40%, 70% et 100%).

Stéphane Géraud est co-animateur dans Aladdin (projet porté par l'INRIA), il a été porteur d'une ACI et il participe à différentes ANR (les pourcentages ne sont pas donnés). Il a aussi été membre de différentes commissions de spécialistes puisqu'il est Maître de conférences, il a été responsable de filière d'enseignement et il partage sa charge d'enseignement entre trois établissements Strasbourgeois depuis septembre 2009. Il a aussi été vacataire à différents endroits comme au CNAM, à l'école d'ingénieurs ENSPS, en DESS : il a donc une expérience d'enseignement riche, variée et envers divers publics.

Analyse du rapport scientifique

Le rapport est particulièrement équilibré et cohérent et il concerne les grilles de calcul et plus particulièrement la programmation par passage de

messages en vue de tirer les performances vers le haut : il s'agit de travaux en calcul haute performance. Le candidat vient du parallélisme « classique » et il a su mettre à profit son expérience pour investir de nouveaux champs. Deux axes de travail sont explorés et résumés dans le document et après la lecture complète du dossier on peut dire que le candidat maîtrise les grands instruments (« à la Grid'5000 »), que sa démarche part souvent d'applications réelles et qu'il a su développer et traiter des thématiques importantes qui sont les suivantes :

1. Applications de géophysique (chapitre 2) : l'objectif visé est clairement de dépasser la situation où le modèle de programmation des clusters est le passage de messages et la situation où la communication dans les grilles se fait selon un paradigme maître-esclave. Il montre que certaines applications doivent être abordées dans le cadre d'un modèle de programmation plus fin et pour le cas des grilles il continue à vouloir travailler par passage de messages. De nombreux problèmes de la géophysique sont abordés, avec du couplage de code. Il est à noter que le candidat s'est tourné vers des applications en dehors du champ informatique et que ce travail a été réalisé en collaboration avec des géophysiciens. Il s'agit donc d'un travail inter-disciplinaire. Stéphane Génaud replace les discussions dans le cadre d'une pile logicielle à adapter à la grille (Fig 2.2) et il montre comment le travail de parallélisation de la brique ray2mesh s'est mis en place. Puis Stéphane Génaud motive le travail autour d'un maillage adaptatif (les capteurs ne sont pas répartis de manière régulière) et donne un algorithme glouton de construction d'un tel maillage irrégulier. Le candidat discute également au chapitre 3 de l'opération collective all-to-all, présente un modèle pour la grille, tire des conclusions pertinentes à partir d'expériences sur de vraies plateformes (il s'agit ici d'une sous partie autour de l'équilibrage de charge statique et la conclusion est attendue dans le sens où il faudrait l'adapter au cas dynamique). La copie correspondante de l'article est donnée dans les annexes. Comme autre perspective du travail pour la problématique de l'équilibrage de charge, mon avis serait d'examiner la technique de résolution quand les fonctions de coûts ne sont plus affines mais convexes. L'hypothèse « affine » permet de transformer le problème en un problème linéaire que l'on résout par des techniques connues. Par contre quand la fonction de coût (du travail réalisé localement sur un nœud n'est plus linéaire) la littérature sur l'équilibrage des charges n'est pas développée à notre connaissance. Un degré supplémentaire de complexité serait de considérer en plus des communications de coût non-affine ;

2. P2P-MPI et applications gridifiées : cette activité liée aux systèmes pairs à pairs qui a pris de l'importance dans la communauté française depuis 4-5 ans est complète dans le sens où elle introduit un nouvel intergiciel avec toutes les propriétés requises, y compris la tolérance aux fautes. De nombreux articles ont été écrits pour ce travail en profondeur qui discutent à la fois des techniques utilisées dans cet outil mais aussi des applications que l'on peut coder, déployer à large échelle avec l'outil. À la lecture des chapitres qui s'y rapportent on mesure que le candidat a développé une activité en architecture des intergiciels jusqu'aux portages d'applications, en passant par la validation et l'écriture de codes non triviaux parce que faisant intervenir différentes techniques de programmation. C'est donc un travail d'ampleur qui est résumé dans le manuscript. A chaque fois, le candidat précise ses hypothèses, son cadre de travail, ce qu'il traite et ce qu'il ne traite pas ce qui rend le document particulièrement lisible.
3. Les chapitres 5 et 6 s'intéressent à des applications parallélisées par le candidat dans le cadre d'un travail pluri-disciplinaire. Le chapitre 5 utilise l'expérience du candidat en matière de Java Message Passing (chapitre 4) et vise à comprendre les limites à l'extensibilité de l'application. Le chapitre 6 s'intéresse à une méthode de classification non-supervisée et de nouveau Stéphane Génaud intervient sur la parallélisation de la méthode en mettant en avant ses compétences sur la gestion des communications sur la grille. L'application permet aussi de « valider » le travail du chapitre 4 (autour de P2P-MPI) en concluant que le couplage de celui-ci avec l'application n'est pas difficile à réaliser.

Perspectives de travail

Le chapitre 7 est dédié à la présentation de pistes de travail. Celles-ci sont organisées selon trois axes :

1. Simulation : ce projet correspondait à un projet récent qui vise en partie à augmenter les capacités de SimGrid pour l'étendre vers la captation de modèles réalistes de passage de messages avec des communications non-bloquantes ; Stéphane Génaud propose également de s'attaquer à la modélisation des communications TCP, non plus dans le cadre d'un cluster (pour lequel les modèles dérivés de LogP peuvent servir) mais dans le cadre de clusters de clusters. Mon avis serait plutôt de se servir de LogP entre les sites en ajoutant du broadcast intra-sites dans le modèle afin de capter les communications multi-sites en général : on en discutera peut-être à la soutenance ;

2. Gestion de données : il s'agit d'étudier la réplication en prenant des informations auprès de l'application (en supposant par exemple que le code source est disponible) ou encore d'ordonnancer des workflows en fonction des données ;
3. Stratégies d'allocation et de découverte de ressources : ce projet correspondant en partie au projet SPADES dans lequel le candidat est impliqué. Il vise à mieux comprendre et à modéliser des machines à base de processeurs multi-coeurs et de rendre les intergiciels plus réactifs. Pour la partie découverte, Stéphane Génaud propose de se concentrer sur les problèmes de connectivité des ressources d'un sous réseau à un autre. Sa conviction est que « l'intergiciel (du futur) devra masquer la complexité des matériels et guider l'utilisateur dans le choix des ressources. »

Conclusion

Au final, Stéphane Génaud, par son manuscript, montre ses qualités d'autonomie scientifique dans plusieurs disciplines du calcul haute performance en contexte de grille. Son travail est bien ancré dans des problématiques importantes pour la communauté : couplage de code intelligent, tolérance aux fautes et P2P pour ne citer que quelques exemples. Son parcours professionnel témoigne également d'une volonté à s'ouvrir à des collaborations extérieures à sa discipline d'origine d'une part et au sein d'institutions variées d'autre part. Sa démarche est exemplaire à bien des égards.

La candidature de Stéphane Génaud mérite donc d'être soutenue et pour toutes les raisons données ci-dessus je suis favorable à une autorisation de soutenance devant le jury proposé.



C. CERIN