

Projet ReMaP

Régularité et parallélisme massif

ENS-Lyon, CNRS & INRIA Rhône-Alpes

THÈME 1A



*R*apport
*d'**A*ctivité

1999

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	5
3	Fondements scientifiques	6
3.1	Compilateurs et bibliothèques	7
3.2	Environnements d'exécution multi-threads	8
3.3	Réseaux haut-débit et calcul coopératif	10
4	Domaines d'applications	11
5	Logiciels	12
5.1	Outils pour le calcul parallèle scientifique	12
5.2	Environnement d'exécution PM2 High-Perf	13
5.3	Services de communications BIP	14
6	Résultats nouveaux	16
6.1	Compilateurs et bibliothèques	16
6.1.1	Transformation de programmes	16
6.1.2	Alignement et distribution des données	17
6.1.3	Algorithmique sur réseau hétérogène de processeurs	18
6.2	Environnements d'exécution parallèles et distribués	19
6.2.1	Actions au niveau de PM2/HighPerf	20
6.2.2	Actions au niveau de Madeleine	21
6.2.3	Support logiciel parallèle hautes performances pour le multimédia	22
6.3	Réseaux haut débit et calcul coopératif	24
6.3.1	Protocoles LAN haut débit	24
6.3.2	Support d'applications parallèles et coopératives	26
6.3.3	Téléphonie mobile	26
7	Contrats industriels (nationaux, européens et internationaux)	26
7.1	<i>LHPC</i>	26
7.2	Collaboration avec la société américaine Myricom	28
8	Actions régionales, nationales et internationales	28
8.1	Actions nationales	28
8.2	Actions européennes	29
8.3	Réseaux et groupes de travail internationaux	30
8.4	Relations bilatérales internationales	30

9	Diffusion de résultats	31
9.1	Animation de la communauté scientifique	31
9.2	Enseignement universitaire	33
9.3	Autres enseignements	34
9.4	Participation à des colloques, séminaires, invitations	35
10	Bibliographie	36

Le projet ReMaP est un projet commun CNRS/ENS Lyon/INRIA du Laboratoire de l'informatique du parallélisme (LIP), UMR CNRS/ENS Lyon/INRIA 5668. Ce projet est localisé à Lyon dans les locaux de l'ENS Lyon.

L'activité des collaborateurs extérieurs se fait en collaboration avec ReMaP, mais se déroule dans le cadre administratif de leurs laboratoires respectifs. La jeune équipe du Ministère RESAM (Université Claude Bernard de Lyon) a été créée au 1er janvier 1999 par Bernard Tourancheau; Laurent Lefèvre et Congduc Pham en sont membres. Ces trois chercheurs ne sont plus associés à ReMaP à compter du 1er septembre 1999: en effet une nouvelle action de l'INRIA Rhône-Alpes, appelée RESO, sera proposée par l'équipe RESAM. Stéphane Ubéda est membre à mi-temps du LIP et à mi-temps du laboratoire TSI, UMR 5516 du CNRS et de l'université Jean Monnet. Récemment nommé Professeur à l'INSA Lyon, il quittera le projet au 1er janvier 2000. ReMaP n'aura donc plus de collaborateur extérieur à cette date.

1 Composition de l'équipe

Responsable scientifique

Yves Robert [Professeur ENS Lyon et UCB Lyon]

Assistants de projet

Sylvie Boyer [Contractuelle, 30% sur le projet]

Anne-Pascale Bottonnet [Adjointe administrative ENS Lyon, 30% sur le projet]

Personnel Inria

Frédéric Desprez [CR]

Jean-François Méhaut [CR, détaché de l'univ. Lille 1 depuis le 01/10/98]

Personnel CNRS

Alain Darté [CR]

Jean-Christophe Mignot [IR]

Loïc Prylli [CR]

Personnel ENS Lyon

Olivier Beaumont [Maître de conférences, arrivée le 01/09/99]

Luc Bougé [Professeur]

Stéphane Domas [ATER à 50%, arrivée le 01/09/99]

Raymond Namyst [Maître de conférences]

Ingénieurs experts

Rémi Choquet [IE Inria, départ le 01/09/99]

Collaborateurs extérieurs

Laurent Lefèvre [Maître de conférences UCB Lyon, départ le 01/09/99]

Congduc Pham [Maître de conférences UCB Lyon, départ le 01/09/99]

Bernard Tourancheau [Professeur UCB Lyon, départ le 01/09/99]

Stéphane Ubéda [Maître de conférences à l'univ. Saint-Étienne, à 50% sur le projet, départ le 01/01/2000]

Chercheurs doctorants

Gabriel Antoniu [Allocataire moniteur MENRT]

Olivier Aumage [Allocataire MENRT, arrivée le 01/09/99]

Alice Bonhomme [Boursière Cifre *MS&I*]

Vincent Boudet [Allocataire moniteur normalien]

Frédérique Chaussumier [Boursière Cifre *MS&I*]

Guillaume Huard [Allocataire MENRT, arrivée le 01/09/98]

Emmanuel Jeannot [Allocataire MENRT, départ le 01/09/99 (Post doc. LaBRI)]

Ahmed Mostefaoui [Boursier algérien MAE]

Christian Perez [Allocataire moniteur normalien]

Cyril Randriamaro [Boursier DRET, départ le 01/09/99 (ATER Amiens)]

Fabrice Rastello [Allocataire MENRT]

Nicolas Schabanel [Allocataire moniteur normalien]

Georges-André Silber [Boursier DRET, départ le 01/09/99 (ATER Marseille)]

Frédéric Suter [Allocataire MENRT, arrivée le 01/09/99]

2 Présentation et objectifs généraux

Mots clés : data-parallélisme, parallélisation automatique, compilation, environnements, bibliothèques, réseaux à haut-débit, applications parallèles.

À sa création, l'objectif du projet *ReMaP* était de contribuer à l'élaboration des connaissances dans le domaine du calcul massivement parallèle régulier. La compilation data-parallèle reste au cœur du projet, mais nous nous intéressons aussi aux applications sur réseaux hétérogènes (bibliothèques, processus légers, protocoles de communication et leurs impacts).

Les axes de recherche de *ReMaP* sont les suivants:

- techniques de parallélisation automatique et outils de parallélisation;
- outils pour l'algèbre linéaire parallèle: algorithmique parallèle en algèbre linéaire dense et creuse, mise en œuvre de bibliothèques numériques sur réseaux hétérogènes de processeurs;
- compilation data-parallèle pour processus légers migrables: HPF (High Performance Fortran) et C*, support d'exécution multi-threads;
- protocoles orientés applications pour les réseaux à haut-débit.

Un point fort du projet est son ancrage industriel et ses activités de transfert. Citons principalement:

- laboratoire *LHPC* commun avec MATRA Systèmes et Information (*MS&I*);
- collaboration avec *MS&I*, le centre multimédia *Érasme*, *Rhône Vision Câble* et *Tonna Informatique* dans le cadre du projet CHARM;
- collaboration avec Myricom autour des développements de BIP;
- responsabilité du TTN *ProHPC* pour le transfert de la technologie du calcul à hautes performances vers les PME;
- coopération avec l'université du Tennessee, Knoxville, dans le cadre de la bibliothèque d'algèbre linéaire parallèle ScaLAPACK qui est mise à la disposition de la communauté internationale.

Une vision schématique mais synthétique de nos axes de recherches est proposée à la Table 1. Les mots-clés de *ReMaP* sont:

Compilation + Bibliothèques + Réseaux + Applications

Niveau	Logiciel	Collaborations	
		France	Internat.
Architectures	Piles de PC, serveur multiméda	<i>MS&I</i>	
Protocoles	BIP, SCI	RHDM, VASY, ARC ResCapA	Myricom, UC Berkeley
Bibliothèques	MPI, PVM, ScaLAPACK		U Tenn., Argonne
Supports d'exécution	PM2, Dosmos	LIFL, Apache, Sirac, Caps	Argonne, Rice
Compilateurs	HPF, C*	ex Paradigme/PRS, iHPerf	GMD, U New Hamp.
Environnements	<i>TransTool, Scilab, Nestor, PlusPyr</i>	Méta2, Résédas, <i>LaBRI</i>	
Applications	Numérique dense + creux Cache Web	<i>LaBRI</i> <i>MS&I</i>	TTN

TAB. 1 – *Vision synthétique de ReMaP*

3 Fondements scientifiques

Depuis quelques années, le paysage des architectures parallèles s'est considérablement modifié, et ce aux deux extrémités du spectre: d'une part le parallélisme est descendu au niveau interne des processeurs (superscalaire, VLIW), et d'autre part on assiste à l'avènement des grappes de PC (*cluster computing*), avec l'interconnexion à plus large échelle de ressources de calcul distribuées (*métacomputing*). Un des nouveaux défis qui se posent pour la mise en œuvre des applications parallèles est la maîtrise de l'hétérogénéité. Pour les utilisateurs et les programmeurs, le caractère hétérogène des architectures pose de nouveaux problèmes dans le développement des applications. Deux approches pour la conception des applications parallèles sont possibles.

- L'application est conçue de manière séquentielle et le programmeur doit pouvoir disposer d'outils de parallélisation automatique et d'interfaces conviviales. Cette problématique est rendue particulièrement complexe avec les nouvelles technologies de processeurs qui permettent par exemple l'exécution simultanée de plusieurs instructions. Les compilateurs de langages, même séquentiels, doivent maintenant réussir à extraire du parallélisme des programmes.
- L'application est conçue de manière parallèle et de nombreuses difficultés surgissent: difficultés algorithmiques bien sûr, mais aussi difficultés de mise au point qui dépendent de la complexité d'une plate-forme parallèle hétérogène. Les principaux problèmes concernent les communications (différentes technologies de réseaux et de protocoles), la répartition des données qui tient compte de l'hétérogénéité (puissance de calcul, mémoire, bande passante) et répartition des calculs.

L'objectif de *ReMaP* est de (tenter de) relever ces nouveaux défis, en apportant des contributions au niveau des algorithmes, des bibliothèques, de la compilation, des environnements de programmation, et des protocoles pour les réseaux à haut-débit. La pyramide des problèmes auxquels nous nous intéressons a été décrite succinctement au paragraphe précédent; ceux-ci seront abordés de manière plus précise ci-dessous.

Notre credo est triple:

- participer aux projets de collaboration internationaux (comme ScaLAPACK) plutôt que de développer en interne des logiciels propriétaires;
- travailler en forte collaboration avec un partenaire industriel pour valider notre approche et nos résultats;
- mener de front des travaux de recherche fondamentale et appliquée.

Ces objectifs, bien que parfois difficilement conciliables, guident nos travaux. Ceux-ci sont structurés en trois axes:

1. compilateurs et bibliothèques;
2. environnement d'exécution multi-threads;
3. réseaux à haut-débit et calcul coopératif.

3.1 Compilateurs et bibliothèques

L'objectif de ce premier thème du projet est de rendre le parallélisme *transparent* pour l'utilisateur, ou du moins de faciliter sa mise en œuvre.

Parallélisation automatique et transformation de programmes Il s'agit de développer et d'intégrer de nouvelles stratégies permettant de transformer (semi-)automatiquement des portions de code séquentiel (principalement des boucles Fortran) en codes annotés par des directives de type HPF (High Performance Fortran) ou OpenMP. Le but est d'aider le programmeur à identifier le parallélisme potentiel de son code au niveau des boucles et d'effectuer automatiquement les transformations nécessaires à sa place (ordonnancement, placement, partitionnement, etc.)

Si les technologies de processeurs évoluent rapidement, elles exigent des compilateurs de plus en plus performants pour les exploiter. Les processeurs actuels disposent aujourd'hui de capacités à exécuter simultanément plusieurs instructions par un pipe-line logiciel. Les techniques de parallélisation automatique étudiées depuis plusieurs années devraient être (ou sont?) utilisées par les prochaines générations de compilateurs. Un des prochains enjeux concernera certainement la synthèse de circuits pour intégrer au niveau du matériel des dispositifs facilitant la conception des compilateurs.

Nos travaux présentent deux facettes, fortement couplées:

- développements théoriques relatifs aux problèmes de transformations de code, de leur classification théorique à la mise au point d'algorithmes les résolvant et à l'étude de la faisabilité de leur intégration au sein d'un compilateur-paralléliseur;

- développements logiciels (*TransTool* et *Nestor*) dont le but est de mettre en œuvre et de valider nos techniques dans le cadre d'une plate-forme de parallélisation (semi-) automatique de programmes Fortran vers HPF. La complexité de la génération du code est un paramètre essentiel pour le choix des transformations source-à-source que nous étudions.

Bibliothèques et algorithmique parallèle hétérogène

Bibliothèque ScaLAPACK: Nous contribuons à plusieurs titres au développement de la bibliothèque d'algèbre linéaire ScaLAPACK pour machines parallèles, qui vise aussi bien la classe des supercalculateurs à mémoire partagée que celle des réseaux de stations de travail à mémoire distribuée:

- optimisation d'enchaînement pipeline de procédures de base d'algèbre linéaire;
- recherche de nouveaux algorithmes de redistribution de tableaux, et amélioration des procédures existantes;
- propositions d'algorithmiques d'extension des noyaux de calcul de ScaLAPACK aux grappes hétérogènes, et aux collections de telles grappes.

Scilab parallèle: *Scilab* est un logiciel de calcul scientifique de type Matlab développé par le projet Méta-2 à l'INRIA. Une intégration de PVM dans *Scilab* permet déjà l'accès à la bibliothèque ScaLAPACK. Nous développons une version parallèle de *Scilab* en collaboration avec Méta-2 et Résédas. Ainsi l'utilisateur pourra-t-il appeler une procédure de la bibliothèque, qui s'exécutera de manière relativement transparente sur les ressources parallèles auxquelles il a accès.

3.2 Environnements d'exécution multi-threads

Environnements multi-threads Depuis quelques années, la notion de processus léger (*thread*) est apparue comme un outil très utile pour le calcul hautes performances. Les threads permettent en effet de recouvrir dynamiquement des tâches de calcul et d'entrées/sorties au sein d'un même nœud de calcul (processus). L'intégration de la notion de thread au sein de C++, puis plus récemment de Java, constitue un très bon exemple de cette évolution. L'étape suivante consiste à utiliser les threads pour le calcul hautes performances parallèle et/ou distribué. Chacun des nœuds du calcul distribué est alors constitué d'un ensemble de threads partageant la mémoire du nœud. Certains threads sont spécialisés dans des tâches de service: gestion des communications avec les autres nœuds, collection d'information de charge, etc. Les autres threads sont dédiés au calcul proprement dit. À cause du très faible surcoût de la gestion des threads, on peut utiliser sans difficulté plusieurs centaines de threads par nœud. Cette approche a déjà été mise en œuvre dans de nombreux systèmes de par le monde, notamment *Nexus* (Ian Foster, ANL) et *Chant* (Matthew Haines, ICASE). En France, on peut citer le système Athapascan 0 (projet Apache, Grenoble), contemporain de PM2.

L'environnement PM2 L'environnement multi-threads distribué PM2 (*Parallel Multithreaded Machine*) a été initialement développé au sein du projet Espace du LIFL, sous la direction de Jean-Marc Geib et de Jean-François Méhaut. Sa principale particularité est de proposer un *noyau de threads* très efficace, ainsi qu'une fonctionnalité de *migration préemptive transparente* des threads entre nœuds distants. Un thread peut à tout moment demander au système la migration de threads présents sur son propre nœud vers des nœuds voisins. La migration est complètement transparente pour le thread migré, du moins dans le cas d'architectures compatibles au niveau du binaire. Ceci laisse la possibilité d'installer au-dessus de PM2 un ordonnanceur global qui équilibre la charge dynamiquement en migrant les threads à travers l'ensemble du système.

Nos axes de recherche À la suite des arrivées de Jean-François Méhaut et Raymond Namy au sein du projet, l'activité de développement de PM2 s'est progressivement déplacée vers Lyon. Notre objectif est de continuer le développement de PM2 en vue des très hautes performances sur des grappes de multiprocesseurs, typiquement des PC interconnectés par des réseaux à très haut débit, par exemple Myrinet, SCI ou bien encore GigaEthernet.

- Au plus bas niveau, cela exige de repenser la structure de PM2 pour l'adapter aux nouveaux protocoles hautes performances qui apparaissent actuellement: protocoles *zéro copie* (BIP, SBP, etc.) et réseaux et protocoles à capacité d'adressage (SCI, VIA, etc.). D'autre part, il est important d'intégrer dans PM2 la notion de *parallélisme hiérarchique* offerte par les grappes de multiprocesseurs.
- Au niveau des processus légers, PM2 est aujourd'hui le seul environnement à s'appuyer sur son propre noyau exécutif. L'avantage d'une telle approche est de pouvoir mieux maîtriser l'ordonnement des processus légers sur les processeurs d'une machine SMP, mais surtout de pouvoir assurer une meilleure collaboration avec le module de communication.
- À un niveau plus élevé, il s'agit d'affiner les politiques d'équilibrage de charge globales pour tenir compte non seulement de la charge de calcul, mais aussi des accès aux données distantes, etc.
- Il s'agit enfin de conduire une mise en œuvre applicative de PM2 comme couche d'exécution de plus gros systèmes, en particulier comme exécutif de compilateurs (HPF Adaptor, C*). L'utilisation de PM2 comme support exécutif pour le langage data-flow Athapascan 1 développé par le projet Apache est actuellement à l'étude.
- Plus généralement, nous étudions l'utilisation des threads pour la construction de serveurs multimédia hautes performances, notamment pour la gestion d'un grand nombre de flux vidéo temps réel parallèles dans le cadre du projet CHARM. Nous avons également étudié une application parallèle de lancer de rayons pour la construction d'images de synthèse réalistes dans le cadre du TTN.

Les performances obtenues sur la plate-forme PoPC (12 PentiumPro 200 MHz, réseau BIP/Myrinet) et sur la plate-forme SCI (8 bi-processeurs Pentium II 450 MHz, réseau SCI)

sont les suivantes :

- commutation entre threads: $< 1 \mu\text{s}$;
- appel à distance d'une procédure vide (*Null RPC*): $10 \mu\text{s}$ sur PoPC, $9 \mu\text{s}$ sur SCI;
- migration d'un thread entre 2 nœuds: $65 \mu\text{s}$ sur PoPC et $25 \mu\text{s}$ sur SCI.

Ces performances nous placent actuellement parmi les meilleurs mondiaux dans ce domaine.

3.3 Réseaux haut-débit et calcul coopératif

L'émergence des réseaux à haut débit est un phénomène nouveau en informatique. La demande des utilisateurs pour le transfert de données complexes et de grande taille (images, vidéo, simulation, etc.) impose de recourir à des technologies matérielles et logicielles nouvelles. Dans ce troisième thème du projet, nous étudions les protocoles des réseaux à haut débit et leurs liens avec les applications coopératives qui les utilisent avec pour but d'en améliorer l'efficacité.

Nos recherches les plus récentes concernent l'étude des réseaux locaux à haut débit ATM et Myrinet et les protocoles associés, ainsi que leurs interactions avec le système d'exploitation. Nos travaux concernent aussi bien le point de vue algorithmique sur ces protocoles, avec la modélisation des architectures réseaux et des constructions algorithmiques pipelines optimisées, que le point de vue pratique pour obtenir des interfaces réseaux et des pilotes de périphériques performants. Notre objectif est de proposer des méthodes et de réaliser des protocoles qui permettent l'utilisation de communication dans les réseaux à haut débit sans (ou presque sans) perte des performances offertes par les matériels du commerce.

Le projet BIP (Basic Interface for Parallelism) a déjà permis d'obtenir de bons résultats en ce sens avec en ensemble de protocoles pour l'utilisation efficace des réseaux Myrinet:

- l'interface *BIP* propose un send/receive de bas niveau avec des performances $< 5 \mu\text{s}$ pour la latence et $> 1 \text{ Gb/s}$ pour le débit;
- le support *MPI-BIP* pour MPICH permet l'utilisation de MPI avec des performances $< 10 \mu\text{s}$ pour la latence et $> 1 \text{ Gb/s}$ pour le débit;
- le pilote *IP-BIP* permet l'utilisation de la pile IP sous Linux avec des performances d'environ $50 \mu\text{s}$ en latence et $> 50 \text{ Moctets/s}$ en débit.

Plus d'une centaine d'équipes utilisent les logiciels BIP comme base pour leurs propres développements (BIP, MPI-BIP) ou comme support réseau performant pour leurs applications (MPI-BIP, IP-BIP). Notons que le pilote IP-BIP est en tête pour les performances TCP et UDP du benchmark *NetPerf* de HP, permettant à des matériels de type PC/Myrinet de rivaliser avec des serveurs spécialisés. Les logiciels BIP sont actuellement disponibles pour l'ensemble des cartes Myrinet et pour les architectures Pentium et Alpha sous Unix (Linux et Solaris) et des travaux d'évaluation sont en cours pour Windows.

Nos axes de recherche concernent les développements autour du support BIP et son évolution. Nous étudions actuellement:

- une nouvelle architecture pour le logiciel *firmware* de BIP,
- une nouvelle politique de gestion de la segmentation et du contrôle de flux dans BIP,
- une interface pour le transport de flux vidéo MPEG,
- les possibilités de support pour le multicast avec les réseaux Myrinet sous BIP,
- le transfert de frame buffer sous BIP,
- l’interfaçage de périphériques directement avec BIP.

Les résultats attendus sont de nouveaux protocoles pour le support des communications rapides et une méthodologie de conception de tels protocoles, grâce à la modélisation des architectures utilisées et la prise en compte des demandes des applications cibles.

Ces recherches interagissent fortement avec le type d’applications pour lesquelles elles sont conduites. Nous avons débuté notre activité *applications coopératives* avec un premier prototype de mémoire virtuellement partagée sur réseaux à haut débit et des travaux concernant la cohérence de données dans les jeux distribués, des applications en imagerie et en mécanique. La prévision du comportement des protocoles débute cette année avec des travaux sur la *simulation parallèle de réseaux* SAN et WAN appliquée à nos protocoles et les utilisant pour l’exécution.

4 Domaines d’applications

Contexte Le parallélisme évolue et les domaines d’application se multiplient. Avec l’apparition de réseaux de stations de travail ou de piles de PC au bon rapport performance/prix, de nombreux utilisateurs se tournent vers des solutions parallèles. Outre les domaines d’application (on dépasse le traditionnel calcul scientifique pour aborder le secteur médical, bancaire, la sidérurgie, le textile, la publicité, la géographie, etc.), la nature des partenaires évolue (aux centres de recherche et développement des grands groupes publics ou privés s’ajoutent désormais les petites et moyennes entreprises). Ces nouveaux utilisateurs ont un besoin crucial d’environnements de programmation performants.

Calcul scientifique et embarqué Nos objectifs se déclinent selon quatre axes principaux.

- Participer en collaboration avec *MS&I* à la réalisation d’une plateforme d’expérimentation en calcul massivement parallèle de haute performance.
- Concevoir et développer un environnement de programmation standard et portable (incluant des bibliothèques de calcul et de macro-communications, des outils de placement et d’ordonnancement, des compilateurs-paralléliseurs, etc.) pour machines parallèles à mémoire distribuée et réseaux de stations de travail.

- Mettre à disposition cette plate-forme d'expérimentation pour les chercheurs et les industriels et offrir des services d'ingénierie pour le portage d'applications existantes ou le développement de nouvelles applications, et pour les activités de recherche associées.
- Assurer la formation des étudiants, chercheurs et ingénieurs et ainsi faciliter l'introduction du calcul parallèle, en particulier au niveau des services de recherche et développement des entreprises.

Applications multimédia Nous développons des logiciels permettant de transformer des grappes d'ordinateurs standards en serveurs de cache Internet à hautes performances pour les têtes de réseau des boucles haut débit. Ce système, réalisé conjointement par *MS&I* et le LIP, utilise des technologies issues du parallélisme pour garantir son extensibilité. Le système de cache intègre des fonctionnalités d'indexation en ligne et de filtrage. Un autre aspect novateur du système réside dans le support des flux audio et vidéo par des mécanismes de cache ou de miroir pour tenir compte de la part grandissante des documents multimédia dans le trafic Internet. Le projet inclut des phases d'expérimentation sur le réseau haut débit *Autoroutes Rhodaniennes de l'Information* de Rhône Vision Câble. Dans le cadre de ces expérimentations, *MS&I* et le LIP mettront un serveur de cache extrêmement performant à disposition du centre serveur Érasme. Cette plate-forme permettra de vérifier l'adéquation des solutions proposées en termes de fonctionnalités et de performances et d'évaluer les bénéfices des infrastructures à haut débit.

Un autre type d'application est étudié dans le cadre du projet SPIHD (Services et Programmes Interactifs pour l'Internet Haut Débit). Plus précisément, il s'agit de définir un ensemble de services de télévision interactive pour l'Internet haut débit. Quatre composantes ont été identifiées pour mener à bien ce projet.

1. La production du contenu multimédia par les partenaires de l'audio-visuel (France3-Lorraine, la cinquième, CanalWeb).
2. Le développement de l'infrastructure logicielle (numérisation, serveur de tête de réseau, gestion et mise à jour des bases de données).
3. L'expérimentation en grandeur réelle sur une boucle locale à Nancy avec la société SEM-Câble de l'Est.
4. L'évaluation technique et économique de la solution pour un éventuel déploiement à plus large échelle.

La contribution du projet *ReMaP* se situe au niveau du point 2 où nous nous intéresserons aux possibilités d'indexation de la base de données des vidéos. Un ingénieur-expert INRIA sera très prochainement engagé pour participer à ce projet et réaliser les développements logiciels.

5 Logiciels

5.1 Outils pour le calcul parallèle scientifique

Mots clés : parallélisation de code, calcul numérique, parallélisation automatique.

Participants : Alain Darte, Frédéric Desprez [correspondant], Guillaume Huard, Cyril Randriamaro, Georges-André Silber.

aLASca aLASca permet la parallélisation d'applications écrites en Fortran en transformant les appels aux bibliothèques de calcul séquentielles BLAS et LAPACK en leur version parallèles ScaLAPACK (et ceci vers Fortran ou HPC). aLASca calcule les distributions de matrices en fonction de la machine cible et tente de supprimer les redistributions inutiles insérées dans le code. aLASca a été intégré à TransTool.

Scilab parallèle *Scilab* est un logiciel de calcul scientifique de type Matlab développé par le projet Méta-2 à l'INRIA. La parallélisation de certains noyaux du logiciel *Scilab* a continué. Nous avons collaboré avec Eric Fleury (projet Résédas, INRIA Lorraine) sur l'interfaçage avec Netsolve, outil de métacomputing développé à l'Université du Tennessee et sur l'interfaçage de la bibliothèque ScaLAPACK. Par ailleurs, nous avons développé un outil de visualisation de matrices creuses distribuées.

Scilab est développé dans le cadre de l'ARC OURAGAN.

Pour plus d'informations voir le site Web: <http://www.ens-lyon.fr/~desprez/OURAGAN/>.

Nestor Le noyau du système de transformations de programmes *Nestor* terminé, ce sont ses fonctionnalités qui ont été enrichies. *Nestor* est un outil de manipulation d'arbres syntaxiques qui gère les entités élémentaires d'un code Fortran comme des objets C++. Ce logiciel (développé par Georges-André Silber), robuste, complètement documenté en HTML, mais simple d'utilisation grâce à son aspect orienté objet, sert de plate-forme d'implantation et de tests pour les algorithmes que nous étudions au niveau théorique. Tous nos algorithmes de transformations de code (voir la section 6.1.1) y sont implantés. Il sert également de noyau à aLaSca. Les nouvelles fonctionnalités de *Nestor* permettent notamment d'éviter la "fuite de mémoire" lors de l'utilisation des objets C++, de disposer de méthodes de parcours de l'arbre syntaxique, avant et arrière, de pouvoir implanter facilement de nouveaux algorithmes par des mécanismes de "plug-in" qui évitent d'avoir à recompiler *Nestor* lors d'un nouvel ajout, etc. Pour plus d'informations voir le site: <http://www.ens-lyon.fr/~gsilber/nestor/>.

5.2 Environnement d'exécution PM2 High-Perf

Mots clés : Multithreading distribué, migration, RPC, réseaux à haut débit.

Participants : Gabriel Antoniu, Olivier Aumage, Luc Bougé, Jean-François Méhaut, Raymond Namyst [correspondant].

PM2 High-Perf est un environnement multithreads portable permettant d'exploiter efficacement les architectures distribuées haute performance (supercalculateurs, grappes de stations SMP interconnectées par réseau haut débit). Il se distingue par l'efficacité de sa gestion des processus légers, par ses fonctionnalités d'équilibrage dynamique de charge (migration de processus légers) ainsi que par son interfaçage efficace avec les protocoles de communication de très bas niveau.

L'objectif de l'environnement PM2 est de définir un cadre de conception et un support d'exécution pour des applications au comportement irrégulier sur architectures distribuées. La caractéristique majeure de ces applications est qu'il est difficile, voire impossible, de répartir statiquement les traitements et les données sur les processeurs de manière équilibrée.

Le modèle de programmation PM2 s'articule autour d'une décomposition des calculs en procédures activables par un mécanisme de type RPC (appel de procédure à distance). Ce découpage, potentiellement extrêmement fin, est pris en charge efficacement par le support exécutif de PM2. Afin de corriger les situations de déséquilibre, PM2 fournit un opérateur de migration permettant de déplacer les activités dynamiquement d'un processeur vers un autre. Cet opérateur s'appuie sur un mécanisme d'allocation iso-adresse assurant qu'un processus léger (et les données qu'il manipule) reste toujours logé dans la même zone d'adresses virtuelles des processus. En complément, PM2 inclut un noyau minimal permettant la gestion d'une partie de l'espace mémoire en mode "virtuellement partagé" (DSM). Ce mécanisme permet aux processus légers d'une application PM2 de partager directement des données en mémoire.

Le support d'exécution s'appuie sur deux bibliothèques (Marcel et Madeleine) qui ont été développées pour PM2. Marcel est une bibliothèque de processus légers qui sont créés en contexte utilisateur (temps de commutation $< 1 \mu s$) et dont l'exécution est prise en charge par des processus noyaux permettant ainsi d'exploiter le parallélisme d'architectures de type SMP. Madeleine est une interface de communication qui se veut à la fois portable et efficace sur différents protocoles réseaux (temps de migration d'un processus léger: $52 \mu s$ sur Myrinet). Récemment, une version adaptative multi-protocoles de Madeleine a été conçue permettant l'utilisation de plusieurs protocoles réseaux au sein d'une même application. En outre, lorsqu'un protocole sous-jacent offre plusieurs modes de transfert des données, Madeleine sélectionne automatiquement le plus approprié de manière à rechercher le maximum d'efficacité. Cette interface est actuellement opérationnelle sur les protocoles TCP, VIA, SCI et SBP.

Les sources complètes de l'environnement PM2/HP sont disponibles à l'URL <http://www.ens-lyon.fr/~rnamyst/pm2.html>. PM2/HP est en outre déposé à l'Agence de Protection des Programmes. Plusieurs équipes de recherches, en France et à l'étranger, utilisent PM2: Lille (Revol), Bordeaux (Roman), Versailles (Roucairol), UNH-Durham/USA (Hatcher), Mons/Belgique (Manneback).

5.3 Services de communications BIP

Mots clés : Passage de messages, communication, réseaux à haut débits, Myrinet.

Participants : Alice Bonhomme, Loïc Prylli [correspondant], Bernard Tourancheau.

Vue d'ensemble BIP est un système logiciel fournissant plusieurs types de services de communications pour le réseau Myrinet.

La couche de plus bas niveau (firmware et bibliothèque hôte) s'interface directement avec le matériel. L'utilisation principale du système BIP se fait par l'intermédiaire de MPI-BIP, une implémentation complète du standard MPI-1 basée sur MPICH, permettant à la majeure partie des applications distribuées de tourner sur notre système sans effort de portage.

Plus de 180 sites distincts ont acquis un mot de passe pour rapatrier le système BIP. Les retours montrent que plusieurs dizaines d'équipes l'utilisent soit pour leurs propres développements de "middleware" (BIP, MPI-BIP), soit comme support réseau performant pour leurs applications (MPI-BIP, IP-BIP).

Les logiciels BIP sont actuellement disponibles pour l'ensemble des cartes Myrinet et pour les architectures x86/Alpha/PowerPc sous Linux,

Nouvelles fonctionnalités bas niveau

- Support système pour passage à un grand nombre de processus par nœud, avec début de mise en place de différentes stratégies d'ordonnancement
- Refonte du système d'enregistrement mémoire pour préserver la sécurité du système: plus de manipulation d'adresses physique dans l'espace utilisateur (dernière étape en cours)
- Support transparent pour l'utilisateur les nouvelles générations de carte Myrinet
- Nouveau protocole de transmissions des petits messages, plus flexible pour permettre par exemple dans le futur:
 - de gérer les pertes de messages et les mécanismes de retransmission dans les situations pathologiques due à un problème matériel,
 - d'offrir un mécanisme de signalisation pour les établissements dynamiques de connexion.

Nouvelle structure du firmware Une expérimentation a été faite pour transformer la structure du code firmware de BIP sous la forme d'un automate de manière à faciliter la réalisation de nouveaux types de protocoles de fragmentation et de contrôle de flot.

Portage NT Un portage partiel sur Windows NT a été effectué, un support plus complet de la partie système de BIP sous NT est envisagé.

Évolutions de MPI Une API intermédiaire a été définie pour porter facilement MPI vers d'autres systèmes et factoriser les parties communes réalisées pour MPI-BIP et MPI-GM, cette trame sert de support pour des portages en cours de notre souche MPI (toujours basée sur MPICH) vers VIA, Madeleine, TCP/IP.

D'autre part notre implémentation de MPI supporte maintenant les configurations hétérogènes.

Composant MPI-SMP Les grappes de SMP permettent de construire des plates-formes parallèle à un meilleur rapport prix par processeur que des solutions uni-processeurs.

Nous avons réalisé un système de messageries pour MPI entre processus du même nœud. Les points forts de ce système sont:

- un composant système permettant les copies directes entres processus sans tampons intermédiaires,

- une intégration facile dans la plupart des implémentations MPI basées sur MPICH pour permettre ou optimiser les communications intra-nœuds (intégration effectuée pour MPI-BIP et MPI-GM),
- des performances supérieures aux implémentations MPI existantes sur les plateformes testées (même par rapport aux implémentations limitées à un seul SMP).

6 Résultats nouveaux

6.1 Compilateurs et bibliothèques

Participants : Olivier Beaumont, Vincent Boudet, Frédérique Chaussumier, Alain Darte, Frédéric Desprez, Stéphane Domas, Guillaume Huard, Emmanuel Jeannot, Cyril Randriamaro, Fabrice Rastello, Yves Robert, Georges-André Silber.

Mots clés : calcul parallèle, parallélisation automatique, environnements de programmation.

Résumé : *Cette partie résume nos travaux récents concernant les compilateurs et les bibliothèques. Ceux-ci sont à la fois théoriques et pratiques: nous développons de nouvelles stratégies de transformation de codes (étude notamment de la fusion de boucles, du pipeline logiciel, des algorithmes de détection de parallélisme), de distribution des données (avant appel à des bibliothèques spécialisées, pour la prise en compte du recouvrement calculs/communications), de nouvelles stratégies algorithmiques (notamment en algèbre linéaire), et également des outils de programmation parallèle (à travers Scilab, TransTool et Nestor).*

6.1.1 Transformation de programmes

À l'heure actuelle, le parallélisme est exploité à tous les niveaux, bien sûr dans les machines clairement identifiées comme parallèles, supercalculateurs, grappes de PCs, mais également au sein même des microprocesseurs, microprocesseurs généraux superscalaires, VLIW (Very Long Instruction Word) et également processeurs dédiés. Pour toutes ces plateformes, aussi différentes soient-elles, l'étude des transformations de programmes révélant du parallélisme et leur automatiser est indispensable.

Algorithme de parallélisation de boucles Nous avons développé un nouvel algorithme de parallélisation de boucles capable, à la différence des algorithmes traditionnels, de traiter de programmes non statiques, c'est-à-dire possédant un flot de contrôle pouvant varier à l'exécution en fonction des données (typiquement des codes avec des IF). Pour cet algorithme, nous avons introduit une nouvelle façon de représenter les dépendances par un graphe appelé "mixed dependence graph" permettant de prendre en compte, simultanément, non seulement les dépendances de données habituelles, mais aussi les dépendances de contrôle. Cette formulation nous permet de contrôler la dimension des tableaux à introduire dans le code final pour

garantir le bon calcul des bornes de boucles et des conditionnelles: l'algorithme résultant qui peut être vu comme une extension de l'algorithme d'Allen, Callahan et Kennedy a été implémenté dans notre outil de transformations de code *Nestor*.

Transformations particulières: fusion de boucles et retiming La fusion de boucles est une transformation de programme qui combine plusieurs boucles en une seule. Elle est utilisée dans les compilateurs-paralléliseurs, principalement pour augmenter la granularité des boucles et pour améliorer la réutilisation des données. Nous avons étudié d'un point de vue théorique plusieurs variantes du problème de fusion de boucles – en identifiant les cas solubles en temps polynômial et les cas NP-complets – et établi le lien entre ces problèmes et quelques problèmes d'ordonnancement provenant de domaines complètement différents.

La transformation dite de *retiming* a de nombreuses applications, à la fois en synthèse d'architectures et en parallélisation automatique. Elle peut être utilisée, en combinaison avec la fusion de boucles, pour augmenter la localité des données, pour diminuer la mémoire nécessaire à un algorithme, pour transformer des boucles séquentielles en boucles parallèles, pour mélanger les itérations d'une boucle et révéler plus de parallélisme (pipeline logiciel), etc. Nous avons proposé de nombreux nouveaux résultats d'optimisation concernant cette technique. Nous avons identifié de nombreux cas surprenants de NP-complétude (tout en proposant des algorithmes de résolution par programmation linéaire en nombre entiers) et d'autres solubles en temps polynômial par des techniques de minimisation de flot dans les graphes. Ces algorithmes ont été implantés dans notre outil de transformations de code *Nestor*.

Synthèse de circuits Nous avons proposé une méthode constructive permettant de résoudre le problème bien connu de l'ordonnancement sans conflits du partitionnement localement séquentiel, globalement séquentiel (LSGP) de la synthèse de réseaux systoliques. Les techniques proposées antérieurement conduisent à des solutions comportant quelques inconvénients pratiques majeurs. Pour y remédier, nous avons donné une expression analytique des solutions qui nous permet d'énumérer tous les ordonnancements valides, ou plutôt qui permet à un générateur automatique de circuits de les énumérer et de les évaluer. Nous avons également développé une technique permettant de réduire le coût du matériel dont le rôle est de contrôler le flot des données, d'activer ou de désactiver les unités fonctionnelles et de générer les adresses mémoire, l'objectif étant de maîtriser la complexité de ces calculs de maintenance du circuit. Ces deux techniques ont été intégrées dans un logiciel de synthèse automatique d'accélérateurs matériels, développé dans les laboratoires de Hewlett-Packard, et ont donné lieu chacune à un brevet.

6.1.2 Alignement et distribution des données

Transformation de code et insertion de redistributions Le logiciel aLaSca, qui permet l'automatisation des transformations d'appels de bibliothèques numériques séquentielles en appels à leur versions parallèles avec calcul des distributions et des redistributions, a été amélioré avec de nouvelles heuristiques qui permettent de traiter de nouveaux cas. Un algorithme mettant en œuvre plusieurs algorithmes du domaine public a été intégré. Le logiciel développé a été intégré à TransTool.

Par ailleurs, nous avons amélioré notre routine de redistributions de matrices en permettant aux messages d'être découpés. Les résultats sont optimaux en nombre de messages envoyés et en nombre d'étapes dans la plupart des cas.

Scilab parallèle Dans le cadre de l'ARC OURAGAN et du logiciel *Scilab*, nous avons travaillé sur l'interfaçage de bibliothèques parallèles comme ScaLAPACK ou PETSc avec l'outil de métacomputing Netsolve. Il s'agit de permettre au programmeur d'avoir différents niveaux d'utilisation du parallélisme: soit le parallélisme est totalement caché grâce à une surcharge des opérateurs d'algèbre linéaire, soit le programmeur a un contrôle sur la distribution des données et des calculs. Deux approches sont retenues, soit par une duplication des processus *Scilab*, soit grâce à un serveur de bibliothèques.

Recouvrements des communications et équilibrage de charge Dans un premier temps, nous avons étudié les recouvrements calculs/communications dans diverses implémentations de la bibliothèque MPI. Nous avons isolé les caractéristiques nécessaires à un recouvrement effectif des communications par les calculs et donné des solutions dans le cas de la bibliothèque MPI-BIP.

Nous avons proposé une nouvelle parallélisation de l'algorithme de rendu volumique du Shear-Warp initialement développé par Lacroute. Nous avons premièrement amélioré l'équilibrage de charge et nous avons proposé des techniques pour le recouvrement des communications par les calculs. Nous avons ainsi obtenu un programme sur grappe de PCs qui peut produire des images de grande taille en temps réel lorsque l'utilisateur change le point de vue.

6.1.3 Algorithmique sur réseau hétérogène de processeurs

Pour distribuer les tableaux d'une application à paralléliser sur un réseau hétérogène de stations de travail, une solution bloc-cyclique n'est pas adaptée. Pour prendre en compte les différentes vitesses des processeurs, on peut penser à une distribution dynamique des données. Mais les stratégies dynamiques peuvent conduire à de mauvais résultats pour deux raisons: (i) le coût des communications liées à d'éventuelles redistributions rendues nécessaires par l'allocation des tâches; (ii) l'inactivité forcée des processeurs à cause des contraintes de dépendances. Nous proposons une approche statique par phases de calcul, en modifiant l'allocation après chacune des phases pour tenir compte d'éventuels changements de vitesse des processeurs. Au sein de chaque phase, nous optimisons la charge des processeurs en fonction de leurs vitesses (ou des estimations courantes de celles-ci).

Ces travaux se sont prolongés dans plusieurs directions au cours de l'année 1999.

Une proposition pour une implémentation de ScaLAPACK sur grappe hétérogène

Nous proposons un algorithme semi-statique, quasi optimal dans le cas où la charge des processeurs ne varie pas, et permettant toutefois des redistributions au vol, de temps en temps, le cas échéant. Nous utilisons des distributions périodiques uni-dimensionnelles des blocs de colonnes des matrices en jeu, et nous avons calculé une distribution optimale pour toute valeur de la période de distribution. Nous avons montré par des tests effectués sur deux plateformes

différentes que cette approche constitue probablement une solution bien adaptée à la parallélisation de plusieurs noyaux de la librairie ScaLAPACK, comme par exemple ceux traitant des décompositions LU ou QR.

Une étude sur grille bi-dimensionnelle hétérogène Nous avons étudié l’implémentation de programmes d’algèbre linéaire tels que la multiplication de matrices ou la résolution de systèmes linéaires, sur une grille hétérogène bi-dimensionnelle de processeurs. Pour ces problèmes, seule une grille 2D assure l’extensibilité des algorithmes utilisés. La distribution classique “bloc-cyclique”, utilisée communément dans le cas d’une grille homogène de processeurs, réduit la performance sur une grille hétérogène à la vitesse du processeur le plus lent. Le problème s’est révélé d’une difficulté surprenante: en effet considérons une grille (virtuelle) fixée de taille $p \times q$; il faut allouer un volume de données inversement proportionnel au temps de cycle de chaque processeur. Si la matrice des temps de cycle n’est pas de rang 1, une telle allocation est impossible et la détermination de la meilleure solution requiert la minimisation d’une fonction quadratique sous des contraintes de même type. Le “vrai” problème n’est pas celui-là, puisque toutes les permutations possibles des processeurs pour définir la grille doivent être envisagées. Nous avons proposé plusieurs heuristiques pour résoudre le problème.

Algorithmique du méta-computing Nos travaux actuels s’orientent vers le problème de l’allocation des données en algèbre linéaire dans le cas de configurations matérielles plus complexes, composées d’une collection distribuée de grappes hétérogènes. Nous avons proposé une première modélisation pour ce type d’architectures et nous étudions actuellement plusieurs techniques d’allocation hiérarchique. Cette étude algorithmique se fait en liaison étroite avec les développements et expérimentations autour de Globus.

6.2 Environnements d’exécution parallèles et distribués

Participants : Gabriel Antoniu, Olivier Aumage, Alice Bonhomme, Luc Bougé, Jean-François Méhaut, Jean-Christophe Mignot, Raymond Namyst, Loïc Prylli, Christian Perez.

Mots clés : calcul hautes performances, multithreading, réseaux hauts débits, PM2, Madeleine, BIP, compilation data-parallèle, migration de threads, serveurs vidéo, applications multimédia.

Résumé : *Les activités dans ce thème ont été cette année consacrées au développement de l’environnement d’exécution PM2/HighPerf de son interface de communication Madeleine et du protocole BIP pour les réseaux Myrinet. Diverses applications dans le domaine de la compilation parallèle, du calcul numérique hautes performances et des entrées-sorties temps-réel pour les serveurs multimédia ont été abordées.*

L’environnement PM2, et tout particulièrement sa couche de communication appelée *Madeleine*, ont été adaptés pour tenir compte de l’évolution des architectures des réseaux et

de protocoles. L'année 1998 avait été consacrée à la mise au point d'une nouvelle version appelée *PM2/HighPerf*, particulièrement optimisée pour les réseaux locaux haut débit. Ceci avait conduit à la refonte complète de la couche de communication, appelée *Madeleine*, digne compagne de la couche de gestion de threads appelée *Marcel!* Cette couche est maintenant constituée d'une interface de portabilité réduite (5 fonctions principales) conçue pour permettre l'exploitation efficace des protocoles de bas-niveau *zéro copie* comme le *BIP*, développé par L. Prylli et B. Tourancheau au-dessus de Myrinet.

En 1999, nous nous sommes concentrés sur le développement de cette couche de communication. Sa conception apparaît comme la clé pour atteindre des performances d'excellente qualité de manière *portable*, c'est-à-dire sur une large classe de protocoles et de matériels, sans modification de l'application. Nous avons notamment étudié la prise en compte des *réseaux à capacité d'adressage* de type SCI dans le cadre de l'ARC *ResCapA*, de manière compatible avec les réseaux plus classiques à échange de messages comme Myrinet ou Giga-Ethernet.

D'autre part, nous avons poursuivi le développement du protocole de communication *BIP* pour compléter son intégration dans les bibliothèques classiques, *MPI* en particulier.

Enfin, nous avons ouvert un nouveau domaine d'activité concernant l'utilisation de ces produits de recherche pour le support des applications multimédia dans le cadre du contrat Jeune équipe CNRS *ParaDigme 2000*, en appui sur le LHPC.

6.2.1 Actions au niveau de PM2/HighPerf

Allocateur iso-adresse pour PM2 Il s'agit de garantir par une stratégie d'allocation dynamique contrôlée globalement que les adresses virtuelles sont préservées lors de la migration des threads. Ceci donne une transparence totale à la migration (entre nœuds compatibles au niveau binaire) et lève toute limitation au niveau de l'utilisation des pointeurs dans les programmes PM2. Ce travail, entamé en 1998 et terminé cette année, a été intégré à la distribution standard cet été. Il constitue le fruit du stage de DEA de Gabriel Antoniu de 1998 soutenu par l'ARC ResCapA.

Support d'exécution pour les compilateurs data-parallèles L'utilisation de PM2 comme *support d'exécution* pour des compilateurs data-parallèles permet de migrer les processeurs virtuels HPF ou C* dynamiquement à l'exécution, de manière transparente pour le compilateur. La version modifiée de la dernière version du compilateur HPF Adaptor (version 6) de Thomas Brandes (GMD) est maintenant opérationnelle, et plusieurs codes HPF ont été testés. Ce travail constitue la contribution centrale de la thèse de Christian Perez. Les bonnes performances obtenues ont conduit et à une évolution dans la conception des prochaines versions de Adaptor.

Mise en œuvre "grandeur nature" de PM2 L'environnement PM2 a servi de base à plusieurs applications. Dans le cadre du TTN *ProHPC*, Raymond Namyst a parallélisé une application commerciale de reconstruction tri-dimensionnelle de meubles de cuisines sur la grappe de PC *PoPC*. L'équipe de Jean Roman au LaBRI, Bordeaux, a utilisé PM2 pour paralléliser des applications hautement irrégulières sur des matrices creuses. Ces expérimentations ont montré l'intérêt de disposer d'une version multiprocesseurs (*SMP*)

de PM2 pour les plateformes bi- ou quadri-processeurs maintenant disponibles. Une version expérimentale SMP de PM2 (en fait de *Marcel*, le noyau de threads de PM2) a été réalisée cet été.

DSM au-dessus de PM2 Nos discussions avec Assaf Schuster (Technion, Haïfa) et Frank Mueller (Humboldt Univ., Berlin) nous ont conduits à envisager d'utiliser PM2 et son iso-allocateur comme outils de base pour un système de gestion mémoire virtuellement partagée (*Distributed Shared Memory, DSM*) spécialisé pour le multi-threading hautes performances et la migration dynamique de threads. Nous nous inspirerons notamment du système *DSMThreads* de Mueller. D'autre part, ce système pourrait constituer la socle de l'environnement de programmation Java distribué *Hyperion* développé par Phil Hatcher (Univ. New Hampshire, Durham, USA). Un premier prototype est en cours de réalisation. L'objectif est comme toujours de maintenir la portabilité des performances sur la plus grande gamme d'architecture de nœuds, de protocoles de communication et de réseaux d'interconnexion possible. Ce travail constituera le cœur de la thèse de Gabriel Antoniu.

6.2.2 Actions au niveau de Madeleine

Toutes ces actions ont été guidées par le souhait d'élargir la portabilité des performances.

Madeleine au-dessus de VIA La norme *Virtual Interface Architecture, VIA* est le résultat d'une concertation entre Compaq, Intel et MicroSoft pour définir une interface de communication hautes performances portable. Cet objectif est donc proche de celui de Madeleine, et nous avons réalisé un portage de Madeleine sur VIA à partir de l'implantation M-VIA sur Fast-Ethernet (Berkeley). Cette implantation de VIA est encore expérimentale, et nous avons eu l'occasion de suggérer un certain nombre d'améliorations qui ont été prises en compte. Les résultats sont cependant intéressants. Même si le succès commercial de VIA est encore incertain, cet essai a montré la possibilité d'exposer au niveau de l'interface Madeleine des fonctionnalités d'adressage distant, en quittant ainsi le modèle des interfaces à base de passage de messages. D'autre part, elle a montré l'importance d'intégrer dans Madeleine une possibilité de choix *dynamique* du protocole utilisé, notamment selon la taille des messages à transmettre.

Madeleine 2 multi-protocoles Cette expérimentation nous a conduits à repenser complètement la structure interne de Madeleine pour permettre à *plusieurs* protocoles, ou plusieurs versions d'un même protocole, d'être gérés en même temps. Le choix du protocole à utiliser peut être imposé par l'utilisateur: par exemple, BIP à l'intérieur d'une grappe, TCP à l'extérieur; il peut être aussi guidé par des heuristiques internes: par exemple pour SCI, procéder par copie synchrone pour les petits messages, et activer le DMA pour les gros. Cette nouvelle structure s'appelle *Madeleine 2* dont une version expérimentale est opérationnelle. Elle est notamment portée sur SCI, et démontre des performances excellentes. Grâce à cette nouvelle version, le temps de migration d'un thread PM2 est passé de 75 μ s sur BIP/Myrinet à 25 μ s sur SISI/DolphinSCI. Ce résultat est à notre connaissance un record du monde! Cette nouvelle conception ouvre la voie à la réalisation

d'un environnement d'exécution hautes performances pour les architectures hétérogènes, grappes de grappes de PC par exemple, dans le cadre d'applications de méta-computing. Des contacts ont été pris avec Carl Kesselman (ISI, Los Angeles) pour coordonner ces travaux avec le développement de l'environnement Globus. La mise au point de Madeleine 2 a constitué l'essentiel du stage de DEA de Olivier Aumage cette année, soutenu par l'ARC ResCapA.

Madeleine sur SCI Un produit direct de cet effort est une implémentation très efficace de Madeleine au-dessus de SCI (en fait, de SISCO), selon nos tâches dans l'ARC *ResCapA*. Nous avons testé avec succès l'environnement PM2 et quelques-unes de nos applications PM2 au-dessus de cette implémentation sur notre nouvelle grappe de PC bi-processeurs SCI. Pour l'instant, nous ne tirons pas parti des bi-processeurs: seul l'un d'entre eux gère les communications. La prochaine étape est bien sûr de lever cette limitation.

Madeleine sur MPC MPC (Multi-PC) est un projet de recherche du LIP6 (Université Pierre et Marie Curie) dirigé par Alain Greiner dont le but est de concevoir une nouvelle technologie de réseau de communication à haut débit pour les grappes de PC. Les éléments originaux de l'architecture MPC sont le réseau d'interconnexion HSL (High Speed Link, standard IEEE 1355) et le dispositif de routage distribué avec les composants de routage Rcube présent sur chacune des cartes d'interface. Trois interfaces de programmation des communications de niveau système (SLRV, SLRP, PUT) sont également développées par l'équipe MPC. Un des objectifs de cette collaboration était d'étudier les problèmes de portage de notre environnement PM2 et de sa couche de communication Madeleine au-dessus de MPC. Ce travail a été réalisé pendant le stage de DEA de Farba Fsy qui a été co-encadré par Philippe Lalevée (INT Evry) et Jean-François Méhaut; une première version de Madeleine sur MPC est aujourd'hui opérationnelle sur la couche SLRV.

6.2.3 Support logiciel parallèle hautes performances pour le multimédia

À partir de l'expérience accumulée dans le domaine des environnements d'exécution multi-threads et des protocoles hautes performances, nous avons ouvert cette année une nouvelle direction de recherche en direction du support des applications multimédia.

Le domaine des serveurs multimédia, actuellement en plein essor, offre une excellente opportunité pour développer notre compétence en système parallèle et en réseau haut débit. Il s'agit par exemple de concevoir des architectures logicielles permettant d'utiliser les grappes de PC pour réaliser des serveurs vidéos ou des caches Web hautes performances à très bas coût en utilisant des composants disponibles en volume dans le commerce. En combinant de tels composants, il est possible d'obtenir des serveurs extensibles jusqu'à de très grandes capacités réseaux, disques et processeurs. Nous sommes convaincus que, après une phase de *développements empiriques*, un certain nombre de *problèmes de recherche fondamentale* vont émerger comme autant de points sur lesquels nous souhaitons apporter une contribution significative:

- technologie de réseaux systèmes (SAN) haut débit;
- utilisation d'interfaces réseaux "intelligentes" hautes performances;

- support logiciel à base de processus légers mobiles pour l'équilibrage de charge;
- systèmes de stockage et d'entrées-sorties adaptées à de très grands volumes d'information avec des contraintes temps réel.

Projet CHARM: Caches hautes performances pour les autoroutes multimédia. Cette nouvelle orientation se matérialise essentiellement par notre participation au projet *CHARM* consacré à l'utilisation des techniques de multithreading et de protocoles hautes performances pour le support des applications multimédia. Ce projet a été présenté au guichet *Autoroutes de l'information* par un consortium dirigé par le Centre Érasme, un établissement de recherche public soutenu par le Conseil général du Rhône et l'Éducation nationale pour étudier l'utilisation des nouvelles technologies multimédia sur réseaux à haut débit à des fins pédagogiques. Le LIP et *MS&I* sont associés dans ce projet pour le développement d'un système de caches Web hautes performances pour des boucles locales haut débit. J.-F. Méhaut et J.-C. Mignot assureront avec L. Prylli la réalisation d'une maquette de cache hiérarchique parallèle faiblement couplée et son optimisation. L. Bougé assure la responsabilité scientifique du projet en ce qui concerne le LIP. A. Bonhomme a été recrutée en thèse sur une bourse Cifre LHPC dans ce cadre. Elle est encadrée par L. Prylli. Pour favoriser cet effort de recherche en liaison avec *MS&I*, nous avons obtenu un financement par le département SPI du CNRS dans le cadre de l'appel à proposition *Junior* du printemps 1999.

Entrées-sorties hautes performances. La thèse de Olivier Aumage démarre cette année sur ce thème. Il s'agit plus précisément d'étudier le problème posé par les entrées-sorties portables et efficaces sur *grappes* de stations (*clusters*). Une des faiblesses des supports exécutifs pour ce type d'architecture est de n'offrir aucune solution simple et satisfaisante aux utilisateurs pour l'accès aux unités de stockages (disques). Le travail consistera donc à étudier et concevoir un support exécutif capable de fournir un accès performant à un volume de données massif sur ces architectures. Le cas des interconnexions de grappes sera aussi étudié. L'approche considérée sera principalement de type *middleware* de manière à maximiser sa portabilité sur différents matériels. Compte tenu des caractéristiques actuelles des grappes de PC en matière de stockage et de communication inter-nœuds (performances réseaux très supérieures aux performances des disques), la mémoire des différents nœuds devra être gérée comme un cache distribué opérant en aval des nécessaires mécanismes d'entrée/sortie disque parallèles. Les données étant de nature non typée, il sera intéressant d'étudier des stratégies adaptatives de redistribution des fichiers sur les disques et d'allocation dans le cache.

Stockage temps-réel. Alice Bonhomme a fait une première année de thèse Cifre consacrée à la conception d'un serveur vidéo réalisé par la société *MS&I*. Sous la direction de L. Prylli, elle s'intéresse au problème du stockage distribué temps-réel et tolérant aux pannes. Ce type de problème a déjà été étudié de manière théorique dans la littérature, mais la réalisation pratique soulève des problèmes peu ou pas abordés:

- concevoir une solution sans contrôle centralisé tout en conservant des clients relativement simples (la tolérance aux pannes reste transparente pour le client);

- gérer les problèmes d'écriture dans les situations de pannes et de récupération après une panne, en conservant le maintien des contraintes temps-réel.

Une synthèse bibliographique sur ce sujet a été rédigée. Elle sera diffusée sous forme de rapport de recherche.

6.3 Réseaux haut débit et calcul coopératif

Mots clés : réseau à haut débit, application coopérative multimédia, protocole SAN, protocole LAN, simulation de réseau, téléphonie mobile, simulation distribuée, objets distribués.

Participants : Alice Bonhomme, Laurent Lefèvre, Congduc Pham, Loïc Prylli, Nicolas Schabanel, Bernard Tourancheau, Stéphane Ubéda.

Glossaire :

mot-clé explication

Résumé : *Cette partie résume nos travaux de l'année 1999 concernant les protocoles pour les réseaux à haut-débit (notamment autour des logiciels BIP). Nous décrivons également nos résultats liés aux réseaux avec nos travaux en algorithmique pour la téléphonie mobile.*

6.3.1 Protocoles LAN haut débit

Les réseaux à haut-débit pour grappe ou LAN continuent à progresser à un rythme rapide. D'une part l'apparition de nouveaux produits ou prototypes (Giganet, ATOLL, MPC) vient renforcer la gamme des technologies existantes (GigaEthernet, Myrinet, SCI, Servernet, Memory Channel). D'autre part le champ d'applications des réseaux Gigabit s'étend vers un public plus large. Ces technologies matérielles permettent de répondre à des besoins croissants des applications: vidéo, images, simulation, calcul scientifique.

Nos recherches récentes concernent l'étude des réseaux locaux Myrinet, GigaEthernet, les protocoles associés ainsi que le type d'intégration entre le système de communication et le noyau du système d'exploitation. Notre objectif est de proposer des méthodes et de réaliser des protocoles qui permettent l'utilisation de communication dans les réseaux à haut débit sans (ou presque sans) perte des performances matérielles offertes par les matériels du commerce.

Le système BIP (comprenant l'API native, le composant MPI, et le pilote IP), dont la conception a commencé en 1997, permet d'obtenir des performances toujours inégalées: performances $< 4 \mu s$ pour la latence et $> 150 Mo/s$ pour le débit, sous BIP, environ $7 \mu s$ sous MPI-BIP. Plus de 180 sites distincts ont acquis un mot de passe pour charger le système BIP, les retours montrent que plusieurs dizaines d'équipes l'utilisent soit pour leurs propres développements de middleware (BIP, MPI-BIP), soit comme support réseau performant pour leurs applications (MPI-BIP, IP-BIP).

Les logiciels BIP sont actuellement disponibles pour l'ensemble des cartes Myrinet et pour les architectures x86/Alpha/PowerPc sous Linux; un portage partiel sur Windows NT a été

effectué, un support complet de la partie système sous NT est envisagé en collaboration avec Microsoft.

Plus précisément, nos activités concernent les points suivants:

- extension du support MPI (en partie via la collaboration avec Myricom) pour convertir la souche MPI développée l'année dernière par nos soins en une interface générique;
- analyse des problèmes de contentions sur le bus PCI, lors du mélange d'opérations de PIO et des DMA, perspectives de modifications des couches basses du pilote TCP/IP;
- analyse des problèmes de recouvrement de calcul-communication dans MPI et des différentes solutions (implémentation multi-threadée, dirigée par interruptions, avec des fonctionnalités "de type "get");
- développement des "channels" associés à l'interface générique développée pour MPI;
- développement d'une interface haute performance pour les transferts au travers d'une mémoire partagée (PC multi-processeurs), compatible avec les échanges réseaux haut débit sous BIP. Ces logiciels "BIP-SMP et SMP-plugin permettent de réaliser des communications entre processus au travers de la mémoire partagée avec un niveau de performance supérieur aux échanges réseaux réalisés avec BIP sous Myrinet. Ce résultat est de première importance pour le développement des grappes de PC multi-processeurs (CLUMPS) et leur utilisation pour des applications à grain fin;
- refonte du noyau de fonctionnement de BIP sur Myrinet avec une méthodologie "event-driven" de manière à améliorer l'extensibilité du logiciel pour les nouvelles fonctionnalités.
- plusieurs collaborations industrielles sont en cours sur ce thème, notamment au travers du LHPC, et avec la société Myricom.

Nos perspectives concernent:

- les développements autour du support BIP et son évolution;
- les nouvelles possibilités de gestion de la segmentation et du contrôle de flux dans BIP;
- les nouvelles fonctionnalités pour permettre les mouvements de données directs entre périphériques (framebuffer, disques);
- les nouvelles possibilités pour le support multi-processus, et les problèmes d'ordonnement/scrutation;
- la réalisation d'une nouvelle interface pour MPI supportant la majorité des supports réseaux existants (BIP, GM, VIA, mémoire partagée, IP, Madeleine);
- la mise en place d'un support pour les flux UDP embarqué sur les cartes GigaEthernet intelligentes.

6.3.2 Support d'applications parallèles et coopératives

Nous avons développé une expérience importante dans les environnements pour l'utilisation et la gestion de grappes de PC. Nous avons montré que, grâce aux réseaux rapides, les grappes étaient maintenant de véritables réponses pour la résolution opérationnelle de nombreux types de problèmes dépassant le cadre bien connu des applications numériques.

Nous avons mis en place une collaboration industrielle avec EDF-DER sur ce thème et nous entretenons une collaboration CEE-PAI avec Linz ainsi que des liens avec l'Université du Tennessee-ORNL.

Nos études sur le support d'applications coopératives donnent une première évaluation des expériences d'ingénierie collaborative effectuées sur des plateformes ATM.

6.3.3 Téléphonie mobile

L'activité sur la planification de réseaux sans fils se poursuit. Dans le domaine des réseaux cellulaires, les développements ont porté sur un démonstrateur écrit en Java des algorithmes d'allocation de fréquence proposés les années précédentes. Un prototype du logiciel existe ; il est en cours de validation. Des perspectives de collaboration existent avec l'équipe du Professeur Tabbanne de l'École Nationale des Postes et Télécommunications de Tunis sur le thème de la prédiction de couverture et l'optimisation dans les réseaux GSM.

Le Databroadcast est une réponse aux problèmes rencontrés dans les réseaux asymétriques (tels serveur/console, WWW, base/téléphones mobiles) où les communications sont plus lentes dans un sens que dans l'autre. Nous avons étudié sa généralisation au cas de plusieurs canaux d'émission pour lequel il possède une heuristique dont il faudra étudier les performances. Nous avons aussi commencé l'étude d' ϵ -approximation avec Neal Young (USA) cet été, lors d'une visite à Dartmouth College. Nous avons publié la version définitive d'une heuristique à facteur constant de l'optimal dans le cas de la diffusion de messages avec préemption dans le cas de messages de longueur non uniforme. Nous avons étudié principalement le problème des messages dont les temps de transmission ne sont pas uniformes, problème que nous avons prouvé NP-dur et pour lequel nous avons proposé une heuristique à facteur constant de l'optimal.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 LHPC

Le *LHPC* (Laboratoire pour les Hautes Performances en Calcul) est un laboratoire commun de recherche sur les ordinateurs massivement parallèles et le calcul à hautes performances créé entre l'INRIA (projet *ReMaP*), l'ENS Lyon, le CNRS et la société MATRA Systèmes & Information (*MS&I*). Il s'appuie sur le contrat de plan état-région avec la Région Rhône-Alpes. Le *LHPC* a été officiellement inauguré le 2 décembre 1996. La convention de collaboration a été renouvelée au 1^{er} janvier 1999, pour une période de quatre ans.

La vocation du *LHPC* est d'être un centre de compétence dans le domaine du calcul parallèle et de réunir des partenaires d'origines variées autour d'un projet académique et industriel aux larges ambitions. Les objectifs du *LHPC* peuvent se résumer selon 4 axes principaux.

- Participer à la réalisation d'une plate-forme de calcul modulaire et extensible.
- Concevoir et développer un environnement de programmation portable mais permettant l'exploitation optimale de la plate-forme d'expérimentation.
- Mettre la plate-forme d'expérimentation à la disposition des chercheurs et des industriels et leur offrir des services d'ingénierie pour le portage d'applications existantes ou le développement de nouvelles applications, ainsi que pour les activités de recherche associées.
- Assurer la formation continue des étudiants, chercheurs et ingénieurs.

Les plate-formes de calcul les plus récentes du *LHPC* sont des piles de PC: pile de Pentium MMX + Myrinet (12 processeurs); pile de Power PC + Myrinet (16 processeurs); pile de Pentium II + SCI (16 processeurs); pile de Pentium II + Myrinet (8 processeurs); pile de Alpha + Myrinet (4 processeurs); serveur multimédia (16 processeurs, 400 Giga-octets de disque). Ces plateformes vont être reliées entre elles par des liens rapides pour permettre des expérimentations algorithmiques, et au niveau système, sur des collections de grappes.

Les principales collaborations au sein du LHPC en 1999 ont été:

- le centre de transfert HPCN *TTN ProHPC*,
- le projet CHARM,
- le montage du projet SPIHD.

Les activités au sein du TTN sont décrites dans le paragraphe 8.2. Les projets CHARM et SPIHD sont résumés ci-dessous.

Projet CHARM Le projet CHARM (*Cache haut débit pour les autoroutes multimédia*) est financé par le SERICS, programme *Autoroutes de l'information*. Il est présenté par un consortium dirigé par le Centre Érasme, un établissement de recherche public soutenu par le Conseil général du Rhône et l'Éducation nationale pour étudier l'utilisation des nouvelles technologies multimédia sur réseaux à haut débit à des fins pédagogiques. Le LIP et *MS&I* sont associés dans ce projet pour le développement d'un système de caches Web hautes performances pour des boucles locales haut débit. J.-F. Méhaut, J.-C. Mignot et L. Prylli assurent la réalisation d'une maquette de cache hiérarchique parallèle faiblement couplée et son optimisation. L. Bougé assure la responsabilité scientifique du projet en ce qui concerne le LIP. Alice Bonhomme (encadrée par L. Prylli) et Frédérique Chaussumier (encadrée par F. Desprez) entament leurs deuxièmes années de thèse sur des bourses Cifre LHPC dans ce cadre. Le projet CHARM se termine en octobre 2000.

Montage du projet SPIHD Le projet SPIHD (*Services et Programmes pour l'Internet Haut Débit*) le projet SPIHD est également mené en collaboration avec *MS&I*, des professionnels de l'audiovisuel (France 3, Canal Web, La Cinquième) et des partenaires de la région lorraine (le LORIA et SEM-Cable de l'Est). L'objectif du projet est de développer une approche de production et de diffusion de contenus multimédias et de journaux télévisés sur l'Internet

haut débit. Les expérimentations seront menées dans Nancy et sa région. Les architectures de grappes de PC développées au LHPC seront utilisées comme serveur de *proxy*. Les tâches de ReMaP incluent des études sur les techniques de code et décodage de vidéo et sur les mécanismes d'indexation à mettre en place pour guider les utilisateurs dans la recherche spécifiques d'information. Un ingénieur-expert INRIA sera embauché à partir du 3 janvier 2000.

7.2 Collaboration avec la société américaine Myricom

La collaboration avec Myricom s'est poursuivie cette année avec un nouveau séjour de L. Prylli dans la société. La prestation fournie à Myricom a portée sur les points suivants:

- introduction de mécanismes d'instrumentation dans GM pour la visualisation et l'analyse de performances au niveau firmware, IP et applications,
- poursuite des travaux MPI-GM, notamment le support zéro-copie, restructuration du code pour minimiser la partie spécifique au système et création d'une API intermédiaire pour un portage très rapide vers d'autres systèmes,
- expertise système Linux, notamment pour le support des nouveaux noyaux et des nouvelles architectures (Sparc, PowerPc), et le support d'architectures mixtes 32bits/64bits.

8 Actions régionales, nationales et internationales

8.1 Actions nationales

Un grand nombre d'actions nationales du projet ReMaP se font dans le cadre du GDR ARP (*Architecture, Réseaux et systèmes, Parallélisme*) dirigé par Michel Diaz (LAAS, Toulouse) avec l'aide de Luc Bougé et Daniel Litaize (IRIT, Toulouse).

ARC ResCapA L. Bougé, R. Namyst, L. Lefèvre et J.-F. Méhaut font partie de l'action de recherche coopérative (ARC) *ResCapA* sur les réseaux à capacité d'adressage de type SCI. Cette action, créée pour 2 ans à l'automne 1997, est dirigée par Thierry Priol, IRISA. Elle réunit les projets CAPS (IRISA), Apache, ReMaP, Sirac (INRIA Rhône-Alpes), et l'équipe Espace de Jean-Marc Geib au LIFL. Elle se traduit notamment au LIP par la mise en place d'une plate-forme d'expérimentation en 1999, constituée d'une grappe de PC bi-processeurs interconnectés par des cartes SCI Dolphin. Cette action a soutenu le stage de DEA de Olivier Aumage et Mickaël Gauthier au sein du projet ReMaP en 1999. Olivier Aumage est maintenant en thèse dans le projet.

ASP MENRT iHPerf'98 En juillet 1997, l'inter-GDR a présenté au MENRT une série de propositions d'Actions sur programme (ASP). Parmi celles-ci, L. Bougé, Jean-Marc Geib (équipe Espace, LIFL, Lille) et Brigitte Plateau (projet Apache, INRIA Rhône-Alpes) ont présenté le projet *iHPerf'98*, une *initiative informatique pour les hautes performances*. (Le texte de ce projet est disponible à l'URL: <http://www.ens-lyon.fr/~bouge/iHPerf/iHPerf.html>). Il s'agissait de stimuler et de coordonner les efforts de *validation applicative* des recherches menées en France ces 10 dernières années dans ce

domaine, en liaison étroite avec des partenaires industriels ou académiques non informatiques. Une douzaine de *micro-actions* sont rassemblées. En novembre 1998, le MENRT a doté ce projet de 350 kF. Les premières micro-actions sont en train de se terminer.

GDR ARP, thème iHPerf Le projet ReMaP participe activement aux activités du groupe thème iHPerf du GDR ARP (Architectures, réseaux et parallélisme) sur l'algorithmique et les outils pour parallélisme dans les applications régulières et irrégulières: L. Bougé, F. Desprez, J.-F. Méhaut, R. Namyst, L. Prylli, Y. Robert. URL: <http://www.prism.uvsq.fr/public/jfcollar/ihperf.html>.

GDR ARP, action Grappes L. Bougé, L. Lefèvre, J.-F. Méhaut, R. Namyst, L. Prylli et B. Tourancheau participent aux activités de cette action dirigée par Jean-Louis Pazat, Projet CAPS et avant-projet PARIS, IRISA. URL: <http://www.irisa.fr/grappes>.

ARC OURAGAN Frédéric Desprez, Stéphane Ubéda, Stéphane Domas, Cyril Randriamaro font partie de l'action de recherche coopérative *OURAGAN* sur les outils pour la parallélisation de grands systèmes numériques. Cette action, créée pour 2 ans à l'hiver 1998, réunit les projets Méta-2 (Rocquencourt), ReMaP, Résédas (INRIA Lorraine) et l'équipe ALIENOR du LaBRI (Université Bordeaux I). Les travaux effectués dans ReMaP concernent la parallélisation du logiciel Scilab et plus particulièrement l'automatisation de l'utilisation de bibliothèques numériques sur plate-formes de métacomputing.

GDR ARP, action TAROT Laurent Lefèvre, Congduc Pham, Nicolas Schabanel, Bernard Tourancheau et Stéphane Ubéda participent au groupe TAROT du GDR ARP sur les télécommunications. ReMaP apporte une compétence algorithmique dans ce groupe de recherche animé par Pierre Fraigniaud du LRI. URL: <http://www.lri.fr/~pierre/TAROT>.

8.2 Actions européennes

TTN ProHPC Le TTN ProHPC est l'un des 20 Centres (Nœuds) de transfert de technologie (TTN) sélectionnés par la Commission européenne. Le principal objectif de ces centres de transfert est d'aider les entreprises, en particulier les PME, à intégrer la technologie *HPCN* (High Performance Computing and Networking) en offrant des ressources matérielles et humaines.

Le LIP (*ReMaP*) est le coordinateur de ce Centre de transfert de technologie dont les trois autres partenaires sont *MS&I*, Simulog et l'IRISA.

En France, le TTN ProHPC:

- organise des campagnes d'information: conférences et formations gratuites, campagnes de presse, serveur Web (<http://www.irisa.fr/ProHPC>);
- supervise plusieurs projets de transfert (aide à la gestion contractuelle, financière et administrative, évaluation des résultats);
- assure la promotion des résultats publics au niveau européen;

- stimule l'élaboration de nouveaux projets (consortium, rédaction de propositions, etc.) dont les partenaires peuvent être issus d'une même région.

Les membres de *ReMaP* sont fortement impliqués dans la coordination du centre et dans les transferts technologiques au sein de plusieurs projets associés au TTN. Plusieurs ingénieurs experts ont été embauchés pour mener à bien ces projets.

La gestion du TTN représente un effort majeur du LIP et de *ReMaP*. La taille du consortium (plus de 50 partenaires) et la nature très diverse des activités de transfert et de dissémination rendent notre tâche très lourde. Les permanents concernés en 1999 sont Karine van Heumen (IR LIP) et Yves Robert pour la coordination, et Frédéric Desprez et Raymond Namyst pour les activités.

La revue finale du TTN a eu lieu début septembre 1999 à Bruxelles. Le bilan final après trois ans de travail s'avère très positif, surtout en ce qui concerne les activités, un peu moins pour les actions de dissémination, largement sous-traitées à des experts mais sources de difficultés. Quant à la coordination du projet complet, c'était sûrement une tâche trop ambitieuse pour les ressources humaines du LIP: nous nous contenterons d'être simples partenaires à la prochaine occasion.

8.3 Réseaux et groupes de travail internationaux

Calcul creux et arithmétique d'intervalle Nous avons participé à la mise en place d'une nouvelle collaboration avec Pierre Manneback, Faculté Polytechnique de Mons, Belgique, et Nathalie Revol, Laboratoire d'Analyse Numérique et d'Optimisation de Lille, et Mohamed Daoudi, Université d'Oujda au Maroc, autour de l'utilisation de PM2 pour des applications scientifiques à haute performance. Le projet DAPPI (Développement d'Algorithmes Parallèles pour des Problèmes Irréguliers) est soutenu par la commission européenne dans le cadre de son programme INCO-DC. Les problèmes irréguliers étudiés dans ce projet concernent les domaines de l'algèbre linéaire creuse et l'optimisation globale par intervalles. F. Desprez et J.-F. Méhaut ont participé en octobre 1999 à l'animation de l'école ParDi '99 (<http://spirou.fpms.ac.be/pardi/index.html>) pour les jeunes chercheurs de l'Afrique du Nord intéressés par ce domaine de recherche. L'environnement PM2 a également été installé sur la grappe de PC du Laboratoire de Recherche en Informatique (LaRI) de l'Université d'Oujda.

Laboratoires de Hewlett Packard, USA Alain Darté collabore avec Bob Rau et Rob Schreiber des laboratoires de HP à Palo Alto sur des problèmes de synthèse automatique de circuits. Cette collaboration a donné lieu à deux brevets sur des techniques de transformation et de génération de code.

8.4 Relations bilatérales internationales

Contrat NSF/INRIA, USA L. Bougé, R. Namyst et J.-F. Méhaut collaborent avec l'équipe de Phil Hatcher et Robert Russell, Dept. Comp. Science, Univ. New Hampshire, Durham, NH, USA, dans le domaine des supports d'exécution pour les compilateurs parallèles. Phil

Hatcher a en particulier développé un compilateur data-parallèle C*, et un environnement d'exécution *distribué* pour Java. Cette collaboration est soutenue par un contrat NSF/INRIA conclu à l'automne 1997 pour 2 ans. Il se termine à la fin de l'année civile 1999.

Contrat UIUC/SPI CNRS, USA L. Bougé, R. Namyst et J.-F. Méhaut collaborent avec l'équipe de Catherine Roucairol et Van-Dat Cung, PRISM, Univ. Versailles Saint-Quentin et l'équipe de Laxmikant V. Kalé, Dept. Comp. Science, Univ. of Illinois at Urbana Champaign (UIUC). Cette collaboration s'intéresse à la mise en œuvre d'algorithmes hautes performances d'optimisation combinatoire au-dessus de PM2 et du système Converse de Kalé. Elle est soutenue par un contrat de collaboration bilatérale conclu entre le département SPI du CNRS et l'UIUC à l'été 1998 pour 2 ans. La bibliothèque d'optimisation (BOB) développée par l'équipe de Versailles est aujourd'hui disponible sur les environnements PM2 et Converse. Des études comparatives d'évaluation de performance vont être menées cette année. La mise en place de liaisons transatlantiques performantes vont aussi nous permettre de mettre le développement d'applications d'optimisation combinatoire à plus large échelle (métacomputing).

Action bilatérale, Tunisie Alain Darté est responsable d'un projet CMCU franco-tunisien qui regroupe en Tunisie la Faculté des sciences de Tunis et l'École nationale des sciences de l'informatique, en France le laboratoire PRISM (Versailles), le LMC (Grenoble) et le LIP. Les relations qui en découlent sont notamment des échanges de chercheurs et l'organisation d'une semaine d'exposés sur le parallélisme en novembre 1999 à Tunis.

Université de Maribor, Slovénie La collaboration bilatérale financée par l'APAPE a été renouvelée en 1999. Stéphane Ubéda en assure la gestion. Des travaux sont en cours en collaboration avec Janez Zerovnik.

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

Responsabilité d'animation

GDR CNRS ARP L. Bougé est le responsable adjoint du Groupement de recherche (GDR) CNRS *Architecture, réseaux et systèmes, parallélisme* (ARP). Ce GDR est dirigé par Michel Diaz, LAAS. Il regroupe une petite centaine d'équipes de recherche française du monde académique. URL: <http://www.ens-lyon.fr/LIP/ARP>.

ASP MENRT iHPerf'98 L. Bougé est responsable avec Jean-Marc Geib et Brigitte Plateau de l'Action de soutien sur programme (ASP) intitulée *Initiative informatique 1998 pour les hautes performances* (iHPerf'98). Cette action a pour but la validation applicative des recherches faites ces dernières années dans le monde académique en informatique sur ce sujet, en étroite collaboration avec des utilisateurs finaux industriels ou académiques non informaticiens. Cette action est financée à hauteur de 350 kF par le MENRT à compter de novembre 1998. URL: <http://www.ens-lyon.fr/~LIP/ARP/iHPerf/iHPerf.html>.

Journées Threads '99 Du 4 au 5 mars 1999, J.-F. Méhaut et R. Namyst ont co-organisé (avec Yves Denneulin) les journées *Threads '99* sur le Campus Saint-Martin d'Hères à Grenoble. Ce workshop national s'inscrivait dans le cadre de l'action iHPerf et avait pour but de faire le point sur les techniques avancées dans le domaine du multithreading (voir http://www-apache.imag.fr/manifestations/journees_threads/). Au cours de ces journées, J.-F. Méhaut et R. Namyst ont effectué trois exposés.

Comités de rédaction, de pilotage et de programme

Luc Bougé est membre du *Steering Committee* de la conférence internationale *Euro-Par* sur le parallélisme depuis 1995. Euro-Par '99 avait lieu cette année à Toulouse, avec près de 400 participants venus du monde entier.

Il est coordinateur du *Comité de pilotage* des Rencontres francophones annuelles du parallélisme (*RenPar*). En 1999, RenPar s'est associé avec 2 autres conférences francophones: SympA 5 (architecture) et CFSE 1 (systèmes d'exploitation) pour organiser un événement commun baptisé *Rennes '99*. Plus de 200 personnes ont été accueillies.

Il a été responsable (*chairman*) du *Topic 11: Parallel Programming: Models, Methods and Languages* de Euro-Par '99 à Toulouse. Il a été membre du Comité de programmes des *11^{es} Rencontres francophones sur le parallélisme* (RenPar'11, Rennes, juin 1999), de *International Conference on Parallel Architectures and Compilation Techniques* (PACT '99, Newport Beach, CA, octobre 1999), *The 3rd Workshop on Runtime Systems for Parallel Programming* (RTSPP'99, San Juan, Puerto Rico, avril 1999), *Workshop on Cluster Computing*, *6th Annual Australasian Conference on Parallel And Real-Time Systems* (PART '99, Melbourne, Australie, novembre 1999), organisée par IEEE CS et TFCC.

Yves Robert fait partie de l'Editorial Board de *Integration, the VLSI Journal* (North Holland) dans la section *Algorithms and Architectures*. Il est membre de l'Editorial Board de *Int. Journal Supercomputer Applications* (MIT Press). Il est l'éditeur européen de *Parallel Processing Letters* (World Scientific Publishing). Y. Robert a récemment été membre des comités de programme suivants: CC'99 (8th International Conference on Compiler Construction), Amsterdam; FDL'99 (Forum on Design Languages), Lyon; STACS'2000 (Symposium on Theoretical Aspects of Computer Science), Lille; ICPP'2000 (International Conference on Parallel Processing), Toronto; PDCS'2000 (Parallel Distributed Computing Symposium), Las Vegas.

Yves Robert a co-organisé avec C. Mongenet et F. Vivien (ICPS Strasbourg) et S. Rajopadhye (IRISA Rennes) le colloque *Compilation et parallélisation automatique* qui s'est tenu à Obernai en octobre 1999. Ce colloque a rassemblé la plupart des équipes françaises actives dans le domaine ainsi que quelques conférenciers invités. Un rapport contenant la liste des interventions et leurs résumés a été rédigé. Ce colloque a été l'occasion de lancer un numéro spécial de la revue TSI (éditions Hermès) sur le thème *Techniques de parallélisation automatique*. Les techniques de compilation-parallélisation automatique ayant atteint une certaine maturité, l'objectif de ce numéro thématique est de proposer un panorama le plus complet possible des recherches récentes dans ce domaine.

Yves Robert a organisé l'évaluation du programme 1A (Parallélisme et architecture) de l'INRIA. De nombreux documents, incluant une présentation générale du thème et les présentations des projets du thème, sont disponibles à l'URL <http://www.ens-lyon.fr/~sboyer/EvaluationINRIA/web-1A.html>.

Frédéric Desprez fait partie du comité de programme du journal "Parallel and Distributed Computing Practices" dont l'éditeur en chef est Marcin Paprzycki (<http://orca.st.us/pdcp/>).

Alain Darte fait partie du comité consultatif (advisory board) de la conférence Euro-Par. Il est membre du comité de rédaction de la revue TSI (Technique et science informatiques). Alain Darte a été rapporteur de A3, avant-projet Universités-CNRS-INRIA.

Alain Darte organise le prochain colloque CPC (Compilers for Parallel Computing) qui a lieu tous les 18 mois depuis 1989: Oxford (1989), Paris (1990), Vienna (1992), Delft (1993), Malaga (1995), Aachen (1996), Vadstena (1998). Ce colloque regroupe des spécialistes essentiellement européens et américains de cette discipline, universitaires ou industriels. Cette huitième édition aura lieu à Aussois du 4 au 7 janvier 2000.

9.2 Enseignement universitaire

Responsabilités d'organisation

DEA d'informatique de Lyon L. Bougé était le responsable de la Formation doctorale d'informatique de Lyon (910036) qui s'est éteinte en septembre 1999. Sur 4 ans, environ 150 diplômes de DEA ont été décernés, et environ 100 thèses ont été soutenues.

École doctorale MathIF et DEA DIF L. Bougé est responsable adjoint de l'*École doctorale de mathématiques et d'informatique fondamentale* (ED MathIF) créée à Lyon à la rentrée 1999. Cette école rassemble le DEA de mathématiques pures, le DEA d'analyse numérique et le DEA d'informatique fondamentale (DEA DIF) qui fait suite au DIL en ce qui concerne le LIP et l'équipe RESAM de B. Tourancheau.

Magistère d'Informatique et Modélisation R. Namyst est responsable de la 2^e année de ce magistère pour l'année scolaire 1999–2000.

Programme européen de l'ENS Lyon L. Bougé est co-responsable de ce programme pour l'informatique. Il s'agit d'envoyer les élèves de MIM 2 (maîtrise) passer 6 mois dans une université européenne, et de recevoir au MIM un nombre équivalent d'élèves européens.

Concours d'entrée à l'ENS Lyon Alain Darte (prenant la suite d'Yves Robert) est responsable de la filière informatique du concours d'entrée de l'École normale supérieure de Lyon.

Enseignement

DEA d'informatique de Lyon En 1998–1999, J.-F. Méhaut y donne un cours intitulé *Conception de support d'exécution parallèle*, Y. Robert y donne un cours intitulé *Techniques de*

parallélisation automatique, B. Tourancheau y donne un cours intitulé *Réseaux et parallélisme*, et Stéphane Ubéda y donne un cours intitulé *Graphes et télécommunications*.

En 1999-2000, plusieurs membres du projet enseignent au nouveau DEA d'informatique fondamentale (DIF): A. Darte (*Ordonnancement et parallélisation automatique*), F. Desprez (*Algorithmique parallèle et parallélisation d'applications*), J.-F. Méhaut et R. Namyst (*Supports d'exécution parallèle et distribuée*), L. Prylli (*Communications dans les réseaux locaux et système d'exploitation*), S. Ubéda (*Optimisation des réseaux sans fil*). URL: <http://www.ens-lyon.fr/DIF/>.

ENSERB Frédéric Desprez a donné un cours en troisième année à l'École Nationale Supérieure d'Électronique et de Radio-Électricité de Bordeaux (ENSERB) dans la filière informatique sur le placement de données et l'équilibrage de charge pour les applications numériques de grande taille.

DESS Réseau de Lyon I Laurent Lefèvre et Congduc Pham sont responsables du DESS. Laurent Lefèvre, Congduc Pham et Bernard Tourancheau y enseignent. Frédéric Desprez a également donné des cours et travaux pratiques sur MPI dans ce DESS.

Ecole d'ingénieurs ISTIL Laurent Lefèvre et Bernard Tourancheau donnent des cours sur les réseaux en 2e et 3e année.

9.3 Autres enseignements

Yves Robert a été nommé Maître de Conférences d'exercice partiel à l'École Polytechnique pour l'année 1999-2000.

Pierre-Louis Curien et Yves Robert ont été les interrogateurs de l'épreuve orale d'informatique commune au concours d'entrée 1999 des trois écoles normales supérieures (Cachan, Lyon et Ulm). Ils ont rédigé un document de travail qui a été diffusé à l'ensemble des professeurs de classes préparatoires concernés.

Alain Darte a été examinateur de l'oral de mathématiques et informatique du concours d'entrée 1999 de l'ENS de Lyon.

Alain Darte a été invité par l'Institute of Information Science (Academia Sinica, Taiwan) à donner une semaine de cours sur ses travaux de recherche. Ces cours ont été annoncés à l'adresse <http://www.iis.sinica.edu.tw/PARALLEL/darte.html> et ont eu lieu en février 1999.

Alain Darte a été invité à donner un tutoriel et un exposé plus pointu en novembre 1999 à Tunis dans le cadre d'une semaine sur le parallélisme.

Frédéric Desprez a donné un cours et une série de travaux pratiques sur MPI, le nouveau "standard" de programmation par passage de messages des machines parallèles à mémoire distribuée dans le cadre du Pôle Scientifique de Modélisation Numérique (PSMN) de Lyon. Le PSMN rassemble un certain nombre de laboratoires utilisateurs du calcul parallèle. La machine cible est un Sun Enterprise 10000 à 32 processeurs.

Frédéric Desprez a été invité à donner un cours à SupElec dans le cadre de la formation continue sur les environnements pour la parallélisation d'applications numériques.

Frédéric Desprez a été invité à donner un cours sur la parallélisation d'applications numériques par passage de messages lors de l'École d'Automne ParDi organisée à Oujda (Maroc) en octobre 1999.

Frédéric Desprez a été invité à donner un tutoriel sur MPI et OpenMP lors de la conférence Euro-Par '99 organisée à Toulouse en septembre 1999.

Stéphane Ubéda a donné plusieurs séminaires à l'École d'Ingénieurs en Télécommunication de Tunis lors d'un séjour d'une semaine en juillet 1999. Plusieurs axes de collaborations ont été définis entre ReMaP et l'équipe du Professeur Sami Tabbanne.

Plusieurs membres du projet ont participé à l'accueil des classes à l'ENS-Lyon lors de la Semaine de la Science 1999. L. Prylli et O. Beaumont ont notamment présenté des exposés *Construire des serveurs de grande taille pour l'internet* et *Le parallélisme: la vie sociale des ordinateurs*.

9.4 Participation à des colloques, séminaires, invitations

Yves Robert a été invité à donner des séminaires au Computer Science Department de UCSB et UCSD (Californie). Il a été invité à donner une conférence "Algorithmic Issues for (Distributed) Heterogeneous Computing", au ERCIM 10th Anniversary Workshop, Amsterdam (Pays-Bas), à la suite du rapport prospectif ERCIM qu'il a rédigé sur ce thème.

Alain Darté a fait deux séminaires sur la complexité de la fusion de boucles au laboratoire Advanced Computing Systems, Kasahara labs (Tokyo) en février 1999 et à UCSD (San Diego) en mars 1999.

Alain Darté et Yves Robert ont participé au colloque *Instruction-Level Parallelism and Parallelizing Compilation* à Dagstuhl (Allemagne) en mai 1999: Alain Darté a présenté les derniers résultats sur le pipeline logiciel décomposé et Yves Robert sur le calcul hétérogène.

Georges-André Silber a fait un exposé à l'Université de Marseille sur la parallélisation automatique par insertions de directives en novembre 1999.

Guillaume Huard a fait un exposé sur le pipeline logiciel décomposé à l'INRIA Rocquencourt en octobre 1999.

Bernard Tourancheau a été invité à présenter BIP lors de la conférence EuroPVM/MPI99.

L. Bougé a donné deux séminaires aux USA lors de visites à la société Myricom, Los Angeles, et à l'ISI, Los Angeles (C. Kesselman).

R. Namyst a donné des séminaires au Labri de Bordeaux et à l'INT Évry sur le thème de l'intégration efficace du multithreading avec les protocoles réseaux haute performance.

J.-F. Méhaut a participé à une table ronde de la conférence RenPar sur les nouvelles architectures Petaflops et le métacomputing. Il a également donné divers séminaires de présentation de PM2 en France (Sophia-Antipolis, Grenoble) et en Irlande à l'Université de Dublin (University College Dublin).

L. Prylli participe à des échanges avec l'équipe d'Alain Greiner au LIP6 pour la réalisation d'un MPI sur la plate-forme MPC. A. Bonhomme et L. Prylli ont effectué des présentations de leurs travaux dans les équipes de R. Muntz à UCLA, Rajive Bagrodia à UCLA, et Shahram Ghandeharizadeh à USC.

10 Bibliographie

Livres et monographies

- [1] A. DARTE, Y. ROBERT, F. VIVIEN, *Scheduling and Automatic Parallelization*, Birkhäuser, 2000, à paraître.
- [2] J. DONGARRA, B. TOURANCHEAU (éditeurs), *Workshop on Clusters and Computational Grids for Scientific Computing*, Special Issue of Int. J. Supercomputers Applications, 1999.
- [3] J. DONGARRA, B. TOURANCHEAU (éditeurs), *Workshop on Clusters and Computational Grids for Scientific Computing, Part I*, Special Issue of Parallel Processing Letters, 1999.

Thèses et habilitations à diriger des recherches

- [4] A. DARTE, *De l'organisation des calculs dans les codes répétitifs*, Habilitation à diriger les recherches, Univ. Claude Bernard, Lyon, LIP, ENS Lyon, décembre 1999, To appear. Written in French.
- [5] M. EXBRAYAT, *Évaluation parallèle de requêtes relationnelles sur réseau de stations – Le système ENKIDU*, Thèse de doctorat, INSA Lyon, Villeurbanne, France, janvier 1999, <ftp://ftp.ens-lyon.fr/pub/users/LIP/mexbraya/publications/Exb1999PhD.ps.gz>.
- [6] E. JEANNOT, *Allocation de graphes de tâches paramétrés et génération de code*, Thèse de doctorat, ENS Lyon, France, LIP, octobre 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/PhD/PhD1999/PhD1999-08.ps.Z>.
- [7] J.-F. MÉHAUT, *Supports d'exécution pour le calcul parallèle et distribué*, Habilitation à diriger les recherches, Univ. Claude Bernard, Lyon, LIP, ENS Lyon, janvier 2000, To appear. Written in French.
- [8] C. PEREZ, *Compilation des langages à parallélisme de données : gestion de l'équilibrage de charge par un exécutif à base de processus légers*, Thèse de doctorat, ENS Lyon, France, LIP, décembre 1999, To appear. Written in French.
- [9] G. SILBER, *Parallélisation automatique par insertion de directives*, Thèse de doctorat, ENS Lyon, France, LIP, décembre 1999.

Articles et chapitres de livre

- [10] C. BARBERET, L. BRUNIE, F. DESPREZ, G. LEBOURGEOIS, R. NAMYST, Y. ROBERT, S. UBEDA, K. VAN HEUMEN, «Technology transfer within the ProHPC at ENS Lyon», *Future Generation Computer Systems (FGCS) 15*, 1999, p. 309–321.
- [11] P. BOULET, J. DONGARRA, F. RASTELLO, Y. ROBERT, F. VIVIEN, «Algorithmic issues on heterogeneous computing platforms», *Parallel Processing Letters 9*, 2, 1999, p. 197–213.
- [12] P. BOULET, J. DONGARRA, Y. ROBERT, F. VIVIEN, «Static tiling for heterogeneous computing platforms», *Parallel Computing 25*, 1999, p. 547–568.
- [13] P.-Y. CALLAND, J. DONGARRA, Y. ROBERT, «Tiling on systems with communication/computation overlap», *Concurrency: Practice and Experience 11*, 3, 1999, p. 139–153.

- [14] F. CAPELLO, D. LITAIZE, J.-F. MÉHAUT, C. MORIN, S. PETITON, D. TRYSTRAM, «Nouveaux challenges pour la prochaine décennie: le Pétaflops (1000 Téraflops)», *Techniques et sciences informatique (TSI)*, 1999, Article invité, numéro spécial RenPar 11. À paraître.
- [15] M. COSNARD, E. JEANNOT, « Compact DAG Representation and Its Dynamic Scheduling », *Journal of Parallel and Distributed Computing* 58, 3, septembre 1999, p. 487–514.
- [16] A. DARTE, Y. ROBERT, F. VIVIEN, *Languages, Compilation Techniques and Run Time Systems for Scalable Parallel Systems: Challenges and Research Focus*, édition D. Agrawal and S. Pande, LNCS, Springer-Verlag, à paraître, ch. Loop parallelization algorithms.
- [17] A. DARTE, « On the Complexity of Loop Fusion », *Special issue of Parallel Computing*, 1999, à paraître.
- [18] A. FERREIRA, C. KENYON, A. RAU-CHAPLIN, S. UBÉDA, « d-Dimensional Range Search on Multicomputers », *Algorithmica* 24, 3/4, 1999, p. 195–208.
- [19] A. FERREIRA, N. SCHABANEL, « A Randomized BSP Algorithm for the Maximal Independent Set Problem », *Parallel Processing Letters (à paraître)*, 1999, Rapport de recherche LIP correspondant: RR1998-19.

Communications à des congrès, colloques, etc.

- [20] G. ANTONIU, L. BOUGÉ, R. NAMYST, C. PEREZ, « Compiling Data-parallel Programs to A Distributed Runtime Environment with Thread Isomigration », in: *The 1999 Intl Conf. on Parallel and Distributed Processing Techniques and Applications (PDPTA '99), Technical Session on parallel and distributed languages: mechanisms, implementations, and tools, 4*, p. 1756–1762, Las Vegas, NV, juin 1999. Invited paper, <http://www.ens-lyon.fr/~bouge/Biblio/Perez/PDPTA99.ps>.
- [21] G. ANTONIU, L. BOUGÉ, R. NAMYST, C. PEREZ, « Compiling Data-parallel Programs to A Distributed Runtime Environment with Thread Isomigration », in: *8th Workshop on Compilers for Parallel Computers (CPC 2000)*, Aussois, France, janvier 2000, <http://www.ens-lyon.fr/~bouge/Biblio/Perez/PDPTA99.ps>.
- [22] G. ANTONIU, L. BOUGÉ, R. NAMYST, « An Efficient and Transparent Thread Migration Scheme in the PM2 Runtime System », in: *Parallel and Distributed Processing. Proc. 3rd Workshop on Runtime Systems for Parallel Programming (RTSPP '99), Lect. Notes in Comp. Science, 1586*, Held in conjunction with IPSP/SPDP 1999. IEEE TCPP and ACM SIGARCH, Springer-Verlag, p. 496–510, San Juan, Puerto Rico, avril 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-08.ps.Z>.
- [23] G. ANTONIU, L. BOUGÉ, C. PEREZ, « Generic load balancing for HPF programs: Application to the Flame Simulation kernel », in: *The 3rd Annual HPF User Group Meeting (HUG '99)*, Redondo Beach, California, août 1999. Official proceedings in preparation, <http://www.ens-lyon.fr/~bouge/Biblio/Perez/AntoniouBougePerezHUG99.ps>.
- [24] G. ANTONIU, C. PEREZ, « Allocation iso-adresse pour une migration préemptive de processus légers: application à un compilateur HPF », in: *Actes des Rencontres francophones du parallélisme (RenPar 11)*, p. 43–48, IRISA, Rennes, France, juin 1999, <http://www.ens-lyon.fr/~bouge/Biblio/Perez/AntoniouPerez-RenPar11.ps>.

-
- [25] G. ANTONIU, C. PEREZ, « Using preemptive thread migration to load-balance data-parallel applications », in : *Euro-Par '99: Parallel Processing, Lect. Notes in Comp. Science, 1685*, Springer-Verlag, p. 117–124, Toulouse, France, août 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-45.ps.Z>.
- [26] V. BOUDET, A. PETITET, F. RASTELLO, Y. ROBERT, « Data allocation strategies for dense linear algebra kernels on heterogeneous two-dimensional grid », in : *Parallel and Distributed Computing and Systems conference (PDCS'99)*, IASTED Press, 1999.
- [27] V. BOUDET, F. RASTELLO, Y. ROBERT, « Algorithmic issues for (distributed) heterogeneous computing platforms », in : *Cluster Computing Technologies, Environments, and Applications (CC-TEA'99)*, R. Buyya, T. Cortes (éditeurs), CSREA Press, 1999.
- [28] V. BOUDET, F. RASTELLO, Y. ROBERT, « A proposal for a heterogeneous cluster ScaLAPACK (dense linear solvers) », in : *International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'99)*, H. R. Arabnia (éditeur), CSREA Press, 1999.
- [29] V. BOUDET, F. RASTELLO, Y. ROBERT, « PVM implementation of heterogeneous ScaLAPACK dense linear solvers », in : *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, J. Dongarra, E. Luque, T. Margalef (éditeurs), LNCS 1697, Springer Verlag, p. 333–340, 1999.
- [30] L. BOUGÉ, J.-F. MÉHAUT, R. NAMYST, L. PRYLLI, « Using the VI Architecture to build distributed, multithreaded runtime systems: a case study », in : *Proc. 2000 ACM Symposium on Applied Computing (SAC 2000)*, ACM Special Interest Group on Applied Computing (SIGAPP), ACM, Villa Olmo, Como, Italy, mars 2000, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-27.ps.Z>.
- [31] L. BOUGÉ, J.-F. MÉHAUT, R. NAMYST, « Efficient Communications in Multithreaded Runtime Systems », in : *Parallel and Distributed Processing. Proc. 3rd Workshop on Runtime Systems for Parallel Programming (RTSPP '99), Lect. Notes in Comp. Science, 1586*, Held in conjunction with IPPS/SPDP 1999. IEEE TCPP and ACM SIGARCH, Springer-Verlag, p. 468–482, San Juan, Puerto Rico, avril 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-10.ps.Z>.
- [32] L. BRUNIE, L. LEFÈVRE, O. REYMANN, « High Performance Distributed Objects for Cluster Computing », in : *International Workshop on Cluster Computing (IWCC '99)*, IEEE Computer Society Press, Melbourne, Australia, décembre 1999.
- [33] F. CHAUSSUMIER, F. DESPREZ, M. LOI, « Efficient Load-Balancing and Communication Overlap in Parallel Shear-Warp Algorithm on a Cluster of PCs », in : *Euro-Par'99 Parallel Computing*, P. Amestoy, P. Berger, M. Daydé, I. Duff, V. Frayssé, L. Giraud, D. Ruiz (éditeurs), *Lecture Notes in Computer Science, 1685*, Springer Verlag, p. 570–577, 1999.
- [34] F. CHAUSSUMIER, F. DESPREZ, L. PRYLLI, « Asynchronous communications in MPI – the BIP/Myrinet Approach », in : *Recent Advances in Parallel Virtual Machine and Message Passing Interface. Proc. 6th European PVM/MPI Users' Group (EuroPVM/MPI '99), Lect. Notes in Comp. Science, 1697*, Springer-Verlag, p. 485–492, septembre 1999.
- [35] F. CHAUSSUMIER, F. DESPREZ, L. PRYLLI, « Asynchronous Communications in MPI - the BIP/Myrinet Approach », in : *Euro PVM/MPI'99, Lecture Notes in Computer Science*, Springer Verlag, 1999. To Appear.

-
- [36] M. COSNARD, E. JEANNOT, T. YANG, «SLC: Symbolic Scheduling for Executing Parameterized Task Graphs on Multiprocessors», in : *International Conference on Parallel Processing (ICPP'99)*, Aizu Wakamatsu, Japan, septembre 1999.
- [37] A. DARTE, G. HUARD, «Loop Shifting for Loop Compaction», in : *Twelfth International Workshop on Languages and Compilers for Parallel Computing, Lecture Notes in Computer Science*, Springer Verlag, août 1999, à paraître.
- [38] A. DARTE, «On the Complexity of Loop Fusion», in : *International Conference on Parallel Architectures and Compilation Techniques (PACT'99)*, octobre 1999, à paraître.
- [39] Y. DENNEULIN, J.-F. MÉHAUT, R. NAMYST, «Customizable thread scheduling directed by priorities», in : *Proc. Workshop on Multithreaded Execution and Compilation (MTEAC 99)*, G. Gao, D. Tullsen (éditeurs), Held in conjunction with 5th Intl Symp. On High Performance Computer Architecture (HPCA-5), IEEE, Orlando, FL, janvier 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-11.ps.Z>.
- [40] F. DESPREZ, E. FLEURY, C. GOMEZ, S. STEER, S. UBÉDA, «Bringing Metacomputing to Scilab», in : *Computer Aided Control System Design (CACSD 99)*, IEEE, Hawaii, USA, août 1999.
- [41] F. DESPREZ, E. FLEURY, L. GRIGORI, «Scilab//: User interactive application and high performances», in : *Third World Multiconference on Systemics, Cybernetics and Informatics (SCI'99) and Fifth International Conference on Information Systems Analysis and Synthesis (ISAS'99)*, IIS, Orlando, USA, août 1999.
- [42] A. FERREIRA, N. SCHABANEL, «A Randomized BSP Algorithm for the Maximal Independent Set Problem», in : *Proc. of the 4th Int. Symp. on Parallel Architectures, Algorithms, and Networks (I-SPAN '99)*, p. 284–289, juin 1999.
- [43] A. FRABOULET, G. HUARD, A. MIGNOTTE, «Loop Alignment for Memory Access Optimization», in : *Twelfth International Symposium on System Synthesis*, 1999.
- [44] P. GEOFFRAY, L. LEFÈVRE, C. D. PHAM, L. PRYLLI, O. REYMANN, B. TOURANCHEAU, R. WESTRELIN, «High-Speed LANs: New Environments for Parallel and Distributed Applications», in : *EuroPar'99, LNCS*, 1685, Springer-Verlag, p. 633–642, Toulouse, France, août 1999.
- [45] P. GEOFFRAY, L. LEFÈVRE, C.-D. PHAM, L. PRYLLI, O. REYMANN, B. TOURANCHEAU, R. WESTRELIN, «High-Speed LANs: New Environments for Parallel and Distributed Applications», in : *Euro-Par '99: Parallel Processing, Lect. Notes in Comp. Science*, 1685, Springer-Verlag, p. 633–642, Toulouse, France, août 1999.
- [46] P. GEOFFRAY, C. D. PHAM, B. TOURANCHEAU, «Exploiting Clusters of Shared Memory Multiprocessors with BIP-SMP: the Parallel Simulation Application», in : *International Conference on Supercomputing (Workshop on Cluster-Based Computing)*, N. P. Carter, S. S. Lumetta (éditeurs), ACM, p. 21–25, Rhodes, Greece, juin 1999.
- [47] P. GEOFFRAY, L. PRYLLI, B. TOURANCHEAU, «BIP-SMP: High Performance message passing over a cluster of commodity SMPs», in : *Supercomputing (SC '99)*, Portland, OR, novembre 1999.
- [48] C. KENYON, N. SCHABANEL, «The data broadcast problem with non-uniform transmission times», in : *Proc. of the 10th Symp. on Discrete Algorithms (SODA '99)*, SIAM, p. 547–556, janvier 1999. Présenté à TAROT en mai 1998 (groupe de travail national d'algorithmique et télécommunication).

-
- [49] A. MOSTEFAOUI, C. PEREZ, L. BRUNIE, « Serveur de séquences audiovisuelles parallèle sur réseau haut débit: concepts et expérimentations », *in: Actes des Rencontres francophones du parallélisme (RenPar 11)*, p. 127–132, IRISA, Rennes, France, juin 1999.
- [50] A. PETITET, H. CASANOVA, J. DONGARRA, Y. ROBERT, R. WHALEY, « Parallel and Distributed Scientific Computing: A Numerical Linear Algebra Problem Solving Environment Designer's Perspective », *in: Handbook on Parallel and Distributed Processing*, J. Blazewicz, K. Ecker, B. Plateau, D. Trystram (éditeurs), Springer Verlag, 1999. Available as LAPACK Working Note 139.
- [51] C. D. PHAM, C. ALBRECHT, « Optimizing Message Aggregation for Parallel Simulation on High Performance Clusters », *in: MASCOT'99*, IEEE, p. 76–83, College Park, MD, USA, octobre 1999.
- [52] C. D. PHAM, R. BAGRODIA, « HLA Support in a Discrete Event Simulation Language », *in: DIS-RT'99*, IEEE, p. 93–100, Greenbelt, MD, USA, octobre 1999.
- [53] C. PHAM, « High Performance Clusters: A Promising Environment for Parallel Discrete Event Simulation », *in: PDPTA '99, III*, CSREA Press, p. 1299–1304, Las Vegas, NV, USA, juin 1999.
- [54] B. PLANQUELLE, J.-F. MÉHAUT, N. REVOL, « Multi-Cluster Approach with PM2 », *in: The 1999 Intl Conf. on Parallel and Distributed Processing Techniques and Applications (PDPTA '99), Technical Session on Cluster Computing Technologies, Environments and Applications, 2*, p. 779–785, Las Vegas, NV, juin 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-23.ps.Z>.
- [55] B. PLANQUELLE, J.-F. MÉHAUT, N. REVOL, « Multi-protocol communications and high speed networks », *in: Euro-Par '99: Parallel Processing, Lect. Notes in Comp. Science, 1685*, Springer-Verlag, p. 139–143, Toulouse, France, août 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-23.ps.Z>.
- [56] B. PLANQUELLE, J.-F. MÉHAUT, « Communications multi-protocoles et réseaux à haut débit », *in: 11es Rencontres francophones du parallélisme (RenPar 11)*, p. 73–78, IRISA, Univ. Rennes 1, Rennes, INRIA, juin 1999.
- [57] L. PRYLLI, B. TOURANCHEAU, R. WESTRELIN, « The design for a high performance MPI implementation on the Myrinet network », *in: EuroPVM/MPI'99, LNCS, 1697*, p. 223–230, Barcelone, Spain, 1999.
- [58] L. PRYLLI, B. TOURANCHEAU, R. WESTRELIN, « An Improved NIC Program for High-Performance MPI », *in: Workshop on Cluster-Based Computing, International Conference on Supercomputing*, ACM, p. 26–30, Rhodes, Greece, juin 1999.
- [59] N. REVOL, Y. DENNEULIN, J.-F. MÉHAUT, B. PLANQUELLE, « Parallelization of continuous verified global optimization », *in: Proc. of the 19th Conference on System Modelling and Optimization*, IFIP TC7, p. 128–131, Cambridge, England, juillet 1999.
- [60] N. SCHABANEL, « The Databroadcast Problem with Preemption », *in: LNCS Proc. of the 17th Symp on Theoretical Aspect of Computer Science (STACS'2000)*, février 2000. A paraître. Rapport de recherche LIP correspondant: RR1999-49 disponible à <http://www.ens-lyon.fr/LIP>.
- [61] G.-A. SILBER, A. DARTE, « The Nestor Library: A Tool for Implementing Fortran Source to Source Transformations », *in: High Performance Computing and Networking (HPCN'99), Lecture Notes in Computer Science, 1593*, Springer Verlag, p. 653–662, avril 1999.

Rapports de recherche et publications internes

- [62] A. BONHOMME, L. PRYLLI, «Choix d'un service de communication pour un serveur vidéo tolérant aux pannes : étude du service GM», *Research Report n° RR1999-13*, LIP, ENS Lyon, Lyon, France, février 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-13.ps.Z>.
- [63] L. BRUNIE, M. EXBRAYAT, A. FLORY, «Parallel Evaluation of Relational Queries on a Network of Workstations», *Research Report n° RR1999-22*, LIP, ENS Lyon, Lyon, France, mars 1999, Also available as Research Report RR-3638, INRIA Rhône-Alpes. Submitted for publication, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-22.ps.Z>.
- [64] F. CHAUSSUMIER, F. DESPREZ, M. LOI, «Efficient Load-balancing and Communication Overlap in Parallel Shear-Warp Algorithm on a Cluster of PCs», *Research Report n° RR1999-28*, LIP, ENS Lyon, France, 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-28.ps.Z>.
- [65] J.-F. MÉHAUT, Y. ROBERT, «Algorithms and Tools for (Distributed) Heterogeneous Computing: A Prospective Report», *Research Report n° RR1999-36*, LIP, ENS Lyon, Lyon, France, août 1999, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-36.ps.Z>.
- [66] J.-C. MIGNOT, «État de l'art des caches Web», *Research Report n° RR1999-52*, LIP, ENS Lyon, Lyon, France, décembre 1999, Rapport de contrat CHARM D1.1.2, <ftp://ftp.ens-lyon.fr/pub/LIP/Rapports/RR/RR1999/RR1999-52.ps.Z>.

Divers

- [67] O. AUMAGE, *Une librairie de communication portable et adaptative pour réseaux haut-débit*, Rapport de stage de DEA, DEA d'informatique de Lyon, Univ. Claude Bernard, Lyon 1, France, juin 1999, http://www.ens-lyon.fr/~bouge/Biblio/Aumage/rapport_DEA.ps.gz.
- [68] P.-L. CURIEN, Y. ROBERT, «Documents de travail pour l'épreuve orale d'Informatique commune aux trois ENS Cachan, Lyon, et Ulm», Disponible sur le Web: http://www.ens-lyon.fr/~yrobert/Oral99_main.ps.gz, 1999.
- [69] A. DARTE, B. R. RAU, R. SCHREIBER, F. VIVIEN, «A Constructive Solution to the Juggling Problem in Systolic Array Synthesis», CASES'99: 2nd International Workshop on Compiler and Architecture Support for Embedded Systems, Washington, DC, octobre 1999, Pas d'actes de conférences.