idris

La lettre de l'IDRIS

Institut du développement et des ressources en informatique scientifique

Février 2012

***ÉDITORIAL** : Denis Girou - IDRIS*

L'année 2011 aura été pour l'IDRIS une année de transition principalement consacrée à la préparation du renouvellement à venir de nos supercalculateurs. Outre les activités correspondant à nos deux missions essentielles, d'une part l'exploitation de notre parc de calculateurs et des machines de service associées et d'autre part le support applicatif de haut niveau apporté à nos utilisateurs, les principaux événements qui ont marqué l'année écoulée ont été :

— des travaux d'infrastructure très conséquents, étendus sur dix-huit mois, permettant de multiplier notre capacité électrique par un facteur 2,5 pour l'exploitation de la prochaine génération de machines (Rafael Medeiros détaille en page 11 les principaux changements effectués) ;

— la préparation, auprès de GENCI, de la procédure d'appel d'offres lancé par celui-ci en octobre dernier pour le renouvellement de nos supercalculateurs, escompté à partir du second semestre de 2012 ;

— l'arrêt à la fin de 2011 de l'exploitation de notre machine vectorielle ;

— le renouvellement de la direction de l'IDRIS au début de l'année 2011 ;

— la fin du projet européen DEISA (*Distributed European Infrastructure for Supercomputing Applications*) en avril dernier, marquant le terme d'un projet de sept années dans lequel l'IDRIS s'était très fortement investi et qu'il avait dirigé durant ses quatre premières années, et dont les activités sont désormais intégrées au sein du projet PRACE (*Partnership for Advanced Computing in Europe*) dans sa deuxième phase d'implémentation qui a démarré il y a quelques mois (PRACE-2IP, septembre 2011 – août 2013) ;

— la poursuite de notre programme soutenu d'actions de formation, pour un total de plus de 50 journées en 2012 et un nombre cumulé de plus de 300 auditeurs, avec en particulier l'introduction d'une nouvelle formation sur les spécificités de la programmation hybride MPI + OpenMP (Pierre-François Lavallée explique pages 4 à 7 pourquoi ce type de programmation va devenir une clé pour l'utilisation des futures générations de supercalculateurs qui comprendront non seulement un nombre très important de nœuds de calcul, mais également un nombre significatif de cœurs à l'intérieur de chacun d'eux, avec la spécificité d'une architecture à deux niveaux suivant un modèle à mémoire distribuée entre les nœuds joint à un modèle à mémoire partagée à l'intérieur de chaque nœud) ;

— le renforcement de nos activités relatives à tous les aspects liés à la sécurité des systèmes d'information, comme en témoignent les articles de Vincent Ribaillier pages 8 à 10 ;

— des actions de communication en cours de redéploiement, avec en particulier la parution prochaine d'une plaquette de présentation, le projet initié d'une refonte complète de notre serveur Web d'ici la fin de l'année et le redémarrage de la publication de notre lettre d'information, dont ce numéro constitue le premier de la nouvelle série (elles continueront à proposer à la fois des informations sur la vie de notre centre, des articles de nature technologique à visée pédagogique et des articles scientifiques illustrant certaines des avancées rendues possibles par l'utilisation de nos ressources, à l'image ici de l'article des pages 2 et 3 rédigé par des chercheurs du CORIA à propos de la combustion turbulente).

C'est donc avec un grand optimisme que nous abordons cette année 2012 qui va ouvrir un nouveau cycle dans la vie de l'IDRIS, pour le plus grand bénéfice attendu de notre communauté d'utilisateurs.

Sommaire

Analyse par simulation directe de l'allumage d'une flamme turbulente dans une machine à compression rapide

Guillaume Lodier, Pascale Domingo & Luc Vervisch
INSA de Rouen & CORIA CNRS, Campus du Madrillet, 76800 Saint-Étienne-du-Rouvray

Les nouveaux concepts de moteurs à combustion interne impliquent un contrôle précis de l'auto inflammation qui suit la compression rapide d'un mélange réactif. Parmi les points qui restent délicats à pleinement maîtriser, plusieurs sont reliés à l'existence d'un grand nombre de scénarios d'allumage, ceci même pour des mélanges réactifs homogènes en composition chimique. L'analyse expérimentale montre en effet que, pour un ensemble de conditions opératoires fixées, les points d'initiation de la combustion présentent une variabilité importante en termes de position dans l'espace et le temps. En sus, de nombreuses incertitudes subsistent concernant l'influence relative du premier point d'allumage sur l'emballement de la combustion et le développement de la flamme turbulente, qui succède à la phase d'initiation de l'allumage.

Pour apporter des éléments de réponse à ces questions, des simulations numériques directes (DNS) ont été réalisées par le laboratoire CORIA. Ces simulations utilisent le logiciel SiTCom [1] (*Simulating Turbulent Combustion*) et le supercalculateur Babel (IBM BG-P) de l'IDRIS. La géométrie et les conditions opératoires sont celles de la machine à compression rapide étudiée à l'Institut Jean Le Rond d'Alembert [2]. Le gaz injecté est homogène en composition et température. Les maillages utilisés contiennent de l'ordre de 70 millions de nœuds de calcul pour une résolution d'environ 20 microns.

Une méthode de frontières immergées permet de représenter le cylindre sur un maillage structuré, où un mélange réactif s'enflamme après une rapide compression réalisée en 29 ms. Les équations de la mécanique des fluides (Navier-Stokes) et de la thermochimie sont résolues dans un formalisme compressible, incluant donc les effets de l'acoustique.

Les questions abordées concernent la genèse de l'allumage :

— Quels sont les paramètres qui conduisent à un allumage global et homogène du mélange réactif, ou à un allumage plus ponctuel, donc localisé ?

— Les premiers points d'allumage sont-ils positionnés au centre des structures tourbillonnaires, ou plutôt entre ces structures turbulentes ?

— Les zones d'allumage primaires favorisent-elles l'allumage en d'autres points et comment contribuent-elles au développement de la flamme turbulente ?

La figure 1 montre les structures tourbillonnaires qui se développent lors de la compression rapide ; des fluctuations de température sont aussi visibles sur cette figure. L'analyse des simulations montre que ces fluctuations proviennent à la fois du transfert thermique dans les régions de proche paroi et d'effets de compressibilité locaux. L'initiation de la combustion étant sensible aux plus petites variations de température, ces fluctuations vont jouer un rôle majeur dans l'allumage du milieu qui est homogène en composition chimique.

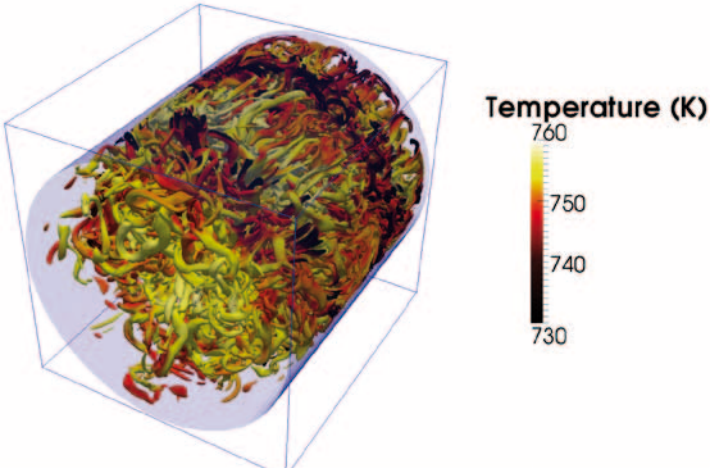


Fig. 1 – Visualisation des structures tourbillonnaires développées lors de la compression rapide

Les simulations sur Babel suggèrent que la distribution de température au sein des structures tourbillonnaires dépend majoritairement de trois phénomènes : la compression adiabatique, l'engouffrement de fluide environnant plus froid en provenance des parois et son mélange au sein des tourbillons. Une compétition s'organise donc entre la compression adiabatique, qui accroît la température des gaz, et le mélange avec le fluide provenant des parois, qui lui décroît localement l'énergie interne du fluide et ainsi la température. Sous ces conditions, deux scénarios d'allumage apparaissent, qui dépendent du temps caractéristique d'allumage piloté par la chimie de l'hydrocarbure utilisé :

— Pour un délai d'allumage plus petit que le temps nécessaire au mélange turbulent pour lisser les fluctuations de température entre les zones proches des parois et le cœur des structures tourbillonnaires, le centre des tourbillons reste thermiquement isolé de son environnement. L'allumage va alors majoritairement se développer à l'intérieur de ces tourbillons, qui sont globalement distribués dans l'écoulement, conduisant à un allumage quasiment en volume, comme observé sur la figure 2 à gauche.

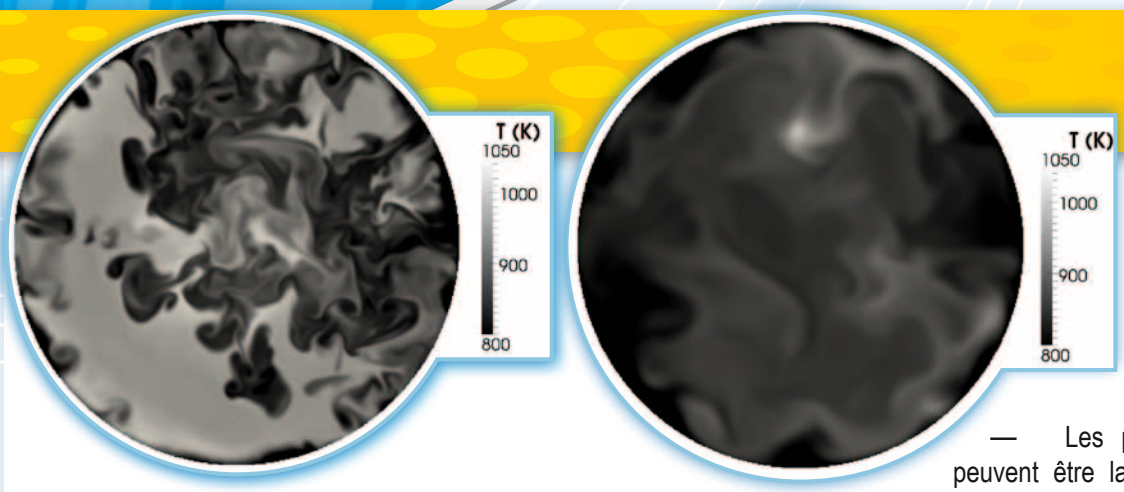


Fig. 2 – Deux régimes d'allumage dans un plan transverse du cylindre : à gauche allumage global sur un demi-cylindre ; à droite allumage localisé en un point

Après l'apparition du ou des premiers noyaux d'allumage, deux chemins conduisent au développement de la flamme :

— Les premiers noyaux de combustion peuvent être la source d'une onde d'allumage, associée à la propagation d'ondes de pression, pour éventuellement favoriser l'allumage lors de leur passage à travers l'écoulement.

— En revanche, si le mélange turbulent entre le fluide proche des parois et le reste de l'écoulement est suffisamment développé, l'allumage va dépendre des moindres détails de la distribution de température, comme par exemple de petits accroissements provenant des zones de compression locales (figure 3), qui apparaissent entre les structures tourbillonnaires. L'allumage présente alors un caractère beaucoup plus ponctuel (figure 2 à droite).

— Dans le cas d'une diffusion efficace de la température dans les zones entourant les structures tourbillonnaires, comme observé sur la figure 4, la propagation d'une onde d'allumage devient moins probable.

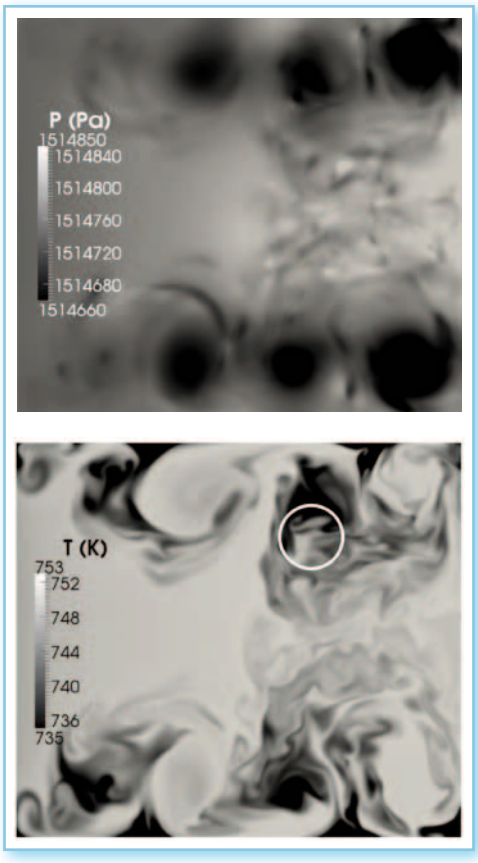


Fig. 3 – Coupe longitudinale du champ de pression (à gauche) et de température (à droite). La zone entourée sur la figure de droite montre une compression entre des structures tourbillonnaires qui conduit au premier point d'allumage

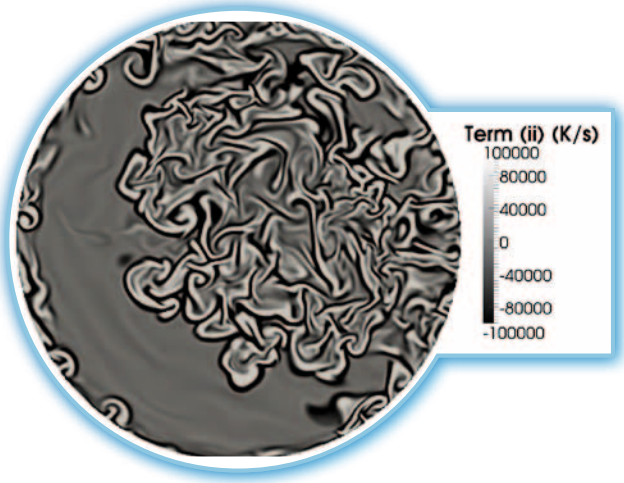


Fig. 4 – Visualisation des gradients de température dans un plan transverse du cylindre

Ces résultats ouvrent des perspectives en termes de contrôle de la séquence d'allumage à partir du contrôle de la turbulence via le procédé d'admission du mélange réactif au sein des cylindres : par exemple, en calibrant le spectre de taille des structures tourbillonnaires, en rapport avec les délais d'allumage du mélange considéré, pour favoriser l'un ou l'autre des deux types d'allumage observés à partir des simulations sur Babel [3].

Références

[1] V. Subramanian, P. Domingo, L. Vervisch (2010) Large-Eddy Simulation of forced ignition of an annular bluff-body burner. Combust. Flame 157(3): 579-601.
[2] P. Guibert, A. Keromnes, G. Legros (2010) An Experimental Investigation of the Turbulence Effect on the Combustion Propagation in a Rapid Compression Machine. Flow, turbulence and combustion, 84(1): 79-95.
[3] G. Lodier, P. Domingo, L. Vervisch Self-ignition scenarios after rapid compression of a turbulent mixture weakly-stratified in temperature. Combust. Flame, submitted.

La programmation parallèle hybride MPI-OpenMP

Pierre-François Lavallée - IDRIS

Introduction

La programmation hybride est un investissement indispensable et pérenne pour passer à l'échelle sur les machines massivement parallèles qui se profilent à l'horizon. Pourquoi est-ce une étape obligatoire ? Quels sont ses inconvénients, ses avantages et ses contraintes ?

Évolution architecturale et avènement du parallélisme massif

L'évolution des architectures de calcul connaît depuis quelques années un changement radical dans la façon dont les gains en performance sont obtenus d'une génération à l'autre. L'évolution de la puissance cumulée des machines du Top500 [1] suit la loi de Moore depuis sa création, soit un doublement de la puissance tous les 18 à 24 mois. Jusqu'aux années 2002, ces gains étaient principalement liés à l'augmentation de la fréquence des processeurs, à une amélioration de la finesse de gravure et à un recours modéré au parallélisme. Aujourd'hui, cette évolution – confortable pour l'utilisateur car elle engendrait peu de changements dans les modèles de programmation – est révolue. La problématique de la consommation électrique et celle du *memory wall* en sont les deux causes principales. La puissance électrique dissipée par un processeur varie comme le cube de sa fréquence : si on double la fréquence, on multiplie par 8 la consommation électrique ! En conséquence, la course à la fréquence est terminée et on observe depuis 2002 une stagnation voire une diminution de la fréquence de fonctionnement des processeurs. Depuis, pour maintenir cet accroissement continu de puissance, il a fallu avoir recours au parallélisme massif à tous les niveaux de l'architecture. Au niveau du cœur d'exécution, on décode plusieurs instructions simultanément (superscalaire), on multiplie et on pipeline les unités fonctionnelles et enfin on a recours à l'*HyperThreading* (HT) ou *Simultaneous Multi Threading (SMT)* pour exécuter simultanément plusieurs threads. Au niveau du processeur, on va multiplier le nombre de cœurs d'exécution et on peut éventuellement lui adjoindre des accélérateurs (GPU, MIC). Au niveau du nœud à mémoire partagée (on utilisera le terme nœud dans la suite), on va multiplier le nombre de processeurs. Enfin, au niveau de la mémoire, on va multiplier le nombre de bancs mémoire accessibles simultanément. Cette (r)évolution s'accompagne d'un changement de référentiel, la sacro-sainte métrique flop/s (la puissance de calcul brute) étant progressivement remplacée parle flop/s/watt (la puissance de calcul par unité électrique) avec son propre classement des machines les plus efficaces énergétiquement, le Green500 [2]. On peut extrapoler l'archétype des machines exaflopiques telles qu'on peut les attendre à horizon 2018-2020. Ce seront des architectures avec des centaines de milliers, voire des millions de cœurs, dont la brique de base sera un nœud potentiellement hétérogène (mélange de cœurs généralistes et de cœurs spécialisés) associé à une hiérarchie mémoire extrêmement complexe composée de multiples niveaux de cache.

Quel modèle de programmation utiliser ?

Si l'évolution des architectures semble plus ou moins tracée à moyen terme, qu'en sera-t-il du modèle de programmation associé et de son influence sur les applications ? Contrairement à ce qui avait été prédit, le passage au Petaflop/s s'est fait sans véritable rupture au niveau du développement des applicatifs, les paradigmes existants s'étant généralement avérés suffisants. Cet état de fait ne sera malheureusement pas de mise pour le passage au multi-Petaflop/s et à l'Exaflop/s.

Les nouveaux langages de type PGAS *(Partitioned Global Address Space)* comme UPC, CAF, Chapel, GA ou X10 sont censés simplifier le développement des applications parallèles via un adressage global de la mémoire qui reste logiquement partitionnée et locale à chaque processeur. Même s'ils ont de véritables qualités, ils ne seront vraisemblablement pas suffisamment matures et performants pour pouvoir être considérés comme une alternative viable à cette échéance. Les modèles existants MPI et OpenMP ont déjà montré leurs limites (approche plate à granularité trop fine pour MPI, limitation à un nœud pour OpenMP) et ne pourront pas être utilisés seuls dans un tel contexte de parallélisme massif. La seule alternative réaliste existe pourtant déjà depuis plusieurs années, c'est la programmation hybride MPI-OpenMP !

Concepts de la programmation hybride

Le concept est simple, consistant à utiliser de façon optimale les modèles existants (MPI et OpenMP) en s'efforçant de s'adapter au mieux à l'architecture précédemment décrite. On va utiliser MPI là où il est indispensable, pour gérer les communications entre les nœuds, et OpenMP à l'intérieur de chaque nœud pour utiliser le plus efficacement possible la mémoire partagée. Concrètement, on va donc avoir recours à un double niveau de parallélisme, chaque processus MPI générant des threads OpenMP qui vont se répartir le travail associé et s'exécuter en parallèle. Le produit du nombre de processus MPI par le nombre de threads OpenMP par processus MPI représente le nombre total de threads d'exécution. Une application « MPI pur » peut alors être vue comme un cas particulier où le nombre de threads par processus MPI est égal à un. Le nombre optimal de threads OpenMP par processus MPI est un paramètre délicat à déterminer ; il est fortement lié aux caractéristiques techniques (réseau, processeur, mémoire, etc.) de l'architecture cible. Cette parallélisation à deux niveaux a de nombreux avantages, mais aussi quelques inconvénients.

Les difficultés de l'approche hybride

L'approche hybride nécessite de gérer deux niveaux de parallélisme imbriqués. C'est indéniablement plus complexe à mettre en œuvre que l'approche MPI seule. Même une simple barrière hybride n'est pas si évidente à implémenter. Elle nécessite rien de moins que l'enchaînement d'une barrière OpenMP, suivie d'une barrière MPI, elle-même suivie d'une seconde barrière OpenMP… Le débogage d'applications hybrides est particulièrement ardu, les outils disponibles étant encore immatures. Concilier portabilité et efficacité est un vrai défi car il existe un nombre important de degrés de liberté à gérer :

— Suivant le niveau de support des threads que l'on utilise, un même algorithme peut s'implémenter de plusieurs façons différentes, parfois de manière très élégante mais totalement inefficace… Choisir la meilleure requiert non seulement une excellente connaissance de l'architecture cible, mais aussi une grande expérience des techniques de programmation et d'implémentation hybride d'applications.
— Au niveau de l'exécution, de nombreux paramètres sont à prendre en compte (le *binding* qui consiste à attacher un thread à un cœur d'exécution, l'affinité mémoire, le rapport optimal du nombre de threads OpenMP par processus MPI, l'optimisation de l'utilisation du réseau d'interconnexion, etc.).

Enfin, la version hybride d'un code ne pourra être efficace que si la version MPI et la version OpenMP le sont. Si un maillon de la chaîne est défaillant, les performances disparaîtront. Mais, fort heureusement, il existe aussi de nombreux avantages à utiliser la programmation hybride, qui contrebalancent largement ces limitations.

L'aspect gain mémoire

Cet aspect est trop souvent passé sous silence lorsque l'on parle de la programmation hybride ; pourtant, c'est l'un de ses avantages essentiels. Le gain va provenir de trois facteurs distincts :

— L'empreinte mémoire des tampons systèmes associés à MPI est non négligeable et croît avec le nombre de processus MPI. Pour donner un ordre d'idée, sur une machine à 65 000 cœurs, l'empreinte mémoire des tampons systèmes peut atteindre 300 Mo par processus, soit plus de 20 To au total !
— Dans les algorithmes de décomposition de domaine, chaque processus MPI doit échanger des informations avec ses voisins, ce qui nécessite l'utilisation artificielle de mailles fantômes (ou *halos*) sur les frontières. Cette problématique disparaît au sein du nœud, puisque les threads ont visibilité sur les données partagées (statut *SHARED*). Dans une approche hybride, seules les mailles fantômes inter-nœuds sont nécessaires. Suivant l'ordre de la méthode de calcul utilisée, le type de domaine (2D ou 3D), le type de décomposition de domaine (mono ou multidimensionnelle) et le nombre de cœurs du nœud, ce gain peut être significatif.
— Dans certaines applications, pour minimiser les communications, on duplique des données sur chaque processus MPI.

Le passage d'une version MPI à une version hybride permet de réduire le nombre de processus MPI (en augmentant le nombre de threads OpenMP), soit implicitement un gain mémoire pour chacun des trois points décrits ci-dessus. Concrètement, sur une version hybride du code CPMD *(Car-Parrinello Molecular Dynamic)* sur la machine CRAY HeCToR de l'EPCC [3], on observe un gain d'un facteur 4,7 par rapport à la version MPI ! Diverses autres application sont également été évaluées dans cette étude, le rapport variant entre 1,8 et 2,9. Très clairement, les gains mémoire obtenus sont significatifs. Cela est particulièrement important dans un contexte général de diminution de la mémoire disponible par cœur d'exécution.

L'aspect dépassement de limites imposées par l'algorithmique

Certaines applications sont parfois limitées en terme d'extensibilité par une taille de domaine, un nombre de zones ou la valeur d'un paramètre physique. C'est par exemple le cas des *NAS Parallel Benchmarks* [4], qui sont des noyaux de calcul extraits d'applications de mécanique des fluides. Plusieurs tailles de problèmes sont disponibles (correspondant à un nombre de zones fixé). Or une contrainte forte de la version parallèle du code MPI est que le nombre de processus ne peut excéder le nombre de zones. C'est extrêmement pénalisant. Sur un problème de classe D par exemple, on a 1024 zones et on est donc limité à 1024 processus MPI ! Grace à l'hybridation du code, tout en gardant le nombre de processus MPI égal à 1024, on peut utiliser plusieurs threads par processus MPI, donc au maximum multiplier le nombre de cœurs d'exécution par le nombre de cœurs du nœud mémoire partagé (soit un facteur 4 sur IBM Blue Gene/P ou un facteur 32 sur IBM Power !).

L'aspect performance et extensibilité

De nombreux facteurs vont permettre d'améliorer les performances et l'extensibilité des codes ayant recours à une parallélisation hybride :

— Meilleure granularité : elle est définie comme le rapport moyen entre deux phases successives de calcul et de communication. L'approche hybride permet d'utiliser le même nombre de cœurs d'exécution, mais avec un nombre de processus MPI réduit. Chaque processus à donc un volume de travail plus important à traiter, ce qui améliore la granularité.

— Meilleur équilibrage de charge (répartition du travail sur les cœurs d'exécution) : un équilibrage de charge dynamique est très consommateur et complexe à implémenter dans une application MPI. Pour une application hybride, au sein d'un processus MPI, un équilibrage de charge dynamique est lui relativement simple à mettre en œuvre, que ce soit via les clauses *DYNAMIC* ou *GUIDED* pour les boucles parallèles OpenMP ou directement à la main en mettant à profit l'accès à la mémoire partagée. Dans un contexte de parallélisme massif, un bon équilibrage de charge (répartition équitable du travail entre les différents processus MPI) est un facteur essentiel pour espérer obtenir une bonne extensibilité d'un code.
— Optimisation des communications : la réduction du nombre de processus MPI minimise le nombre de communications et augmente la taille des messages. L'impact de la latence est ainsi réduit alors que les débits des communications sont améliorés. Cela est encore plus significatif sur les communications collectives dont les coûts sont non linéaires et fonction du nombre de processus MPI mis en jeu.
— Amélioration de la rapidité de convergence de certains algorithmes itératifs, qui utilisent les informations du domaine local à chaque processus MPI. Recourir à moins de processus MPI revient à avoir des domaines locaux de plus grande taille, donc des informations locales plus « riches », ce qui accélère la vitesse de convergence de la méthode numérique (i.e. méthodes itératives à base de préconditionneurs partiels).
— Optimisation des entrées-sorties : la réduction du nombre de processus MPI engendre moins d'accès simultanés en entrées-sorties et augmente la taille moyenne des enregistrements sur disque. Les serveurs de méta-données sont moins chargés et les requêtes de taille plus adaptée au système.
— Approche parfaitement adaptée aux architectures de nouvelle génération qui demanderont une utilisation poussée de l'*hyperthreading* ou du *multi-threading* pour tirer parti de la performance des cœurs de calcul.

Les gains potentiels sont d'autant plus importants que le nombre de cœurs d'exécution est élevé. Si la parallélisation hybride est correctement réalisée, la limite d'extensibilité de la version hybride du code comparée à celle de la version MPI se trouve augmentée d'un facteur pouvant aller jusqu'au nombre de cœurs du nœud…

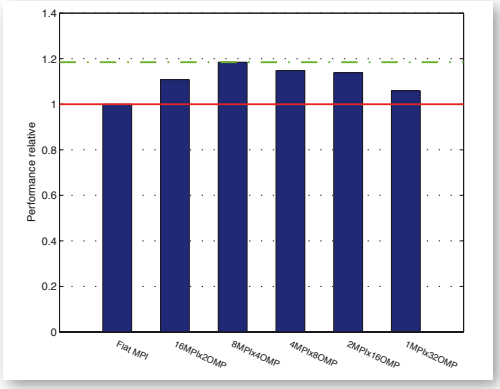
Résultats obtenus sur un code d'hydrodynamique

À l'IDRIS, nous avons travaillé sur l'application d'hydrodynamique HYDRO (méthode volumes finis associée à une méthode de Godunov qui à chaque interface résout un problème de Riemann sur une grille régulière 2D). Ce code a été parallélisé avec MPI via une méthode de décomposition de domaine suivant les deux dimensions du problème. Ensuite, une version OpenMP basée elle aussi sur une méthode de décomposition de domaine a été implémentée à partir de la version initiale. Une fois ces deux versions validées, la version hybride, fusion des deux versions précédentes a été développée en injectant dans la version OpenMP les appels MPI. En procédant de cette façon, les coûts de développement de la version hybride ont été réduits (de l'ordre de quelques jours de travail pour le développement et la validation du code hybride) et finalement le gros du travail aura été le développement des deux versions parallèles initiales (MPI et OpenMP).

Cet exemple est sur le papier un cas particulièrement défavorable à la programmation hybride, le code MPI ayant déjà une excellente extensibilité et un équilibrage de charge quasi-parfait. Y a-t-il vraiment un intérêt à passer à une version hybride dans ce cas ?

Dans un premier temps, nous avons validé notre approche hybride sur un nombre très limité de cœurs d'exécution (64 cœurs, soit deux nœuds de la machine IBM SP6 de l'IDRIS). Pour ce test, on a fait varier le nombre de threads OpenMP par processus MPI de 1 à 32. Les résultats sont donnés dans les tableaux suivants (on a normalisé à 1 la performance de la version MPI) :

MPI x OpenMP par nœud	Temps en s. sur 64 cœurs de calcul
32 x 1	66,8
16 x 2	60,3
8 x 4	56,4
4 x 8	58,2
2 x 16	58,7
1 x 32	63,1

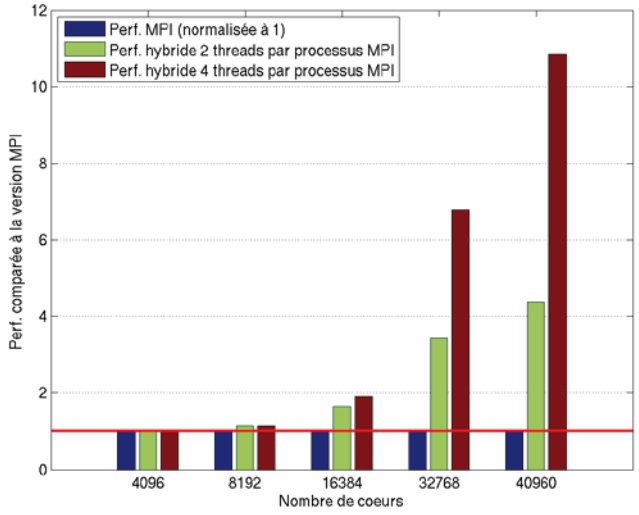


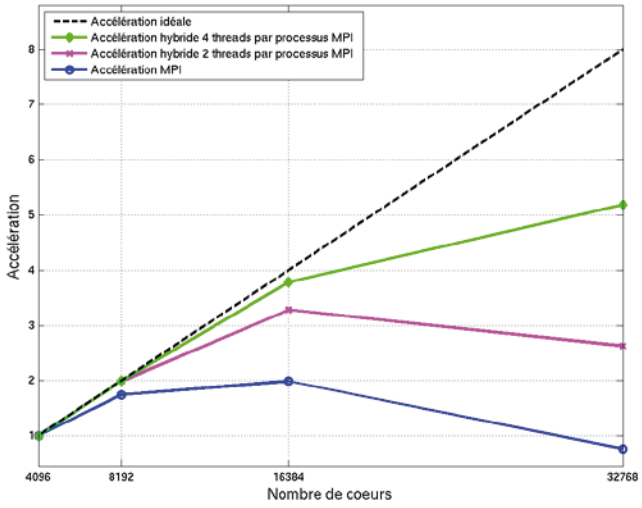
Les conclusions sont les suivantes :

— Les versions hybrides sont toujours plus performantes que la version MPI.
— Le gain maximum est de l'ordre de 20 % pour une répartition sur l'ensemble des deux nœuds de 16 processus MPI et 4 threads OpenMP par processus MPI.
— Même sur un nombre de cœurs limité, l'approche hybride peut donc amener des gains en performance.

Dans un deuxième test, on s'est s'intéressé au parallélisme massif en réalisant sur l'IBM BG/P de l'IDRIS un test d'extensibilité de type *Strong Scaling* (taille totale du domaine fixe quel que soit le nombre de cœurs utilisé). On fait varier le nombre de cœurs d'exécution de 4 096 à 40 960 pour un domaine de 40 000 x 40 000 points de grille. Les résultats sont donnés dans le tableau suivant :

Tps en s.	4 096 cœurs	8 192 cœurs	16 384 cœurs	32 768 cœurs	40 960 cœurs
MPI	61,7	35,4	31,0	80,7	136,7
Hybride 4 threads par proc. MPI	62,4	31,0	16,3	12,0	12,6





Les conclusions sont les suivantes :

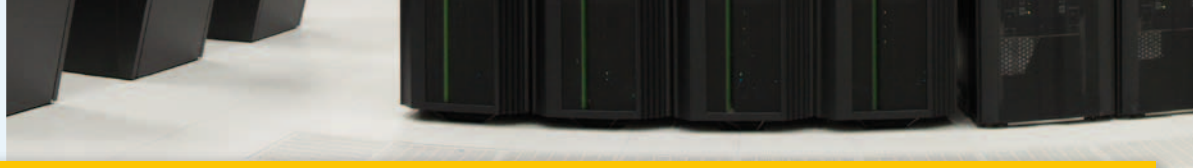
- Sur 4 096 cœurs, la version MPI et la version hybride sont presque équivalentes.
- Sur le domaine considéré, l'extensibilité de la version MPI montre rapidement ses limites, peinant à accroître ses performances jusqu'à 8 192 cœurs, puis s'effondrant.
- Sur le domaine considéré, la version hybride avec 4 threads par processus MPI accroît linéairement ses performances jusqu'à 16 384 cœurs et même jusqu'à 32 768 cœurs, mais plus de façon linéaire. Au-delà, on n'améliore plus les temps de restitution.
- Pour ce test difficile, la limite d'extensibilité de la version MPI est de 8 192 cœurs alors que celle de la version hybride est de 32 768 cœurs. On retrouve ici le facteur 4 qui correspond au nombre de cœurs d'un nœud de la BG/P !
- La meilleure version hydride (sur 32 768 cœurs) est 2,6 fois plus rapide que la meilleure version MPI (sur 8 192 cœurs).

Conclusions

Toutes les applications peuvent bénéficier des avantages de la programmation hybride (gain en mémoire, en performance et en extensibilité, etc.). Cependant, si votre code ne rencontre pas aujourd'hui de problème en termes de limitation mémoire par processus MPI, s'il est parfaitement extensible et que l'équilibrage charge est bon, il ne sera d'aucune utilité d'avoir recours à une version hybride. Mais si vous n'êtes pas dans ce cas idéal et que votre problématique scientifique demande toujours plus de ressources de calcul au fur et à mesure que celles-ci sont disponibles, alors le parallélisme massif et la programmation hybride seront un passage obligé pour espérer tirer parti de la puissance des machines de nouvelle génération. C'est un investissement important mais pérenne. Pour vous accompagner dans cette démarche, l'IDRIS propose depuis l'année dernière une formation dédiée [5] de 4 jours. C'est l'occasion d'approfondir les points abordés dans cet article et de mettre en application les concepts sous-jacents sur des exemples concrets. Tous les résultats présentés dans cet article sont extraits du support de cours hybride qui est librement téléchargeable sur le site Web de l'IDRIS [6].

Références

[1] Top500. [Online]. www.top500.org
[2] Green500. [Online]. www.green500.org
[3] Anastasios Stathopoulos, "Mixed Mode Programming on HECToR", EPCC, MSc in HighPerformance Computing.
[4] NAS Parallel Benchmarks. [Online]. http://www.nas.nasa.gov/publications/npb.html
[5] Pierre-François Lavallée et Philippe Wautelet. (2011) Formation IDRIS – Programmation hybride MPI-OpenMP. [Online]. http://www.idris.fr/data/cours/hybride/choix_doc.html
[6] Site Web de l'IDRIS. [Online]. www.idris.fr



Protection du potentiel scientifique et technique

Denis Girou et Vincent Ribaillier - IDRIS

La *protection du potentiel scientifique et technique* (PPST) de la nation est une préoccupation primordiale qui a toujours animé les gouvernements successifs de la République. Mais face aux changements apparus progressivement ces dernières décennies dans les risques, les menaces et les moyens d'attaque mis en œuvre, il était devenu indispensable de modifier la législation en vigueur en la matière. Bien qu'il ait été amendé plusieurs fois au fil du temps, le dispositif précédent reposait sur une vision relevant pour l'essentiel des rapports de force géopolitiques tels qu'ils existaient à l'époque de la guerre froide. Si ce temps est évidemment révolu, de nouvelles menaces ont vu le jour à partir d'autres zones géographiques et en même temps de nouveaux domaines sont visés, notamment ceux de l'économie et des systèmes d'information pour la protection desquels l'ancienne réglementation était mal adaptée. C'est pour cela qu'après une longue gestation étendue sur deux années, un nouveau dispositif a été établi, basé sur le décret n° 3011-1425 du 2 novembre 2011, dont la mise en application est effective depuis le 1er février 2012.

Conçue en partenariat constant avec des représentants du monde de la recherche, cette nouvelle réglementation vise non pas à freiner mais à favoriser les échanges scientifiques, éléments déterminants de la diffusion des savoirs, en responsabilisant toutefois tous les acteurs, en homogénéisant les espaces de coopération, en diffusant de bonnes pratiques, en écartant les « prédateurs » éventuels et en sanctionnant pénalement les agresseurs. Quatre types de risques majeurs ont été retenus : les atteintes aux intérêts économiques, les atteintes aux capacités de défense, la prolifération d'armes de destruction massive et le terrorisme.

La nouvelle réglementation s'étend dorénavant *erga omnes*, c'est-à-dire à tous et non plus comme par le passé uniquement aux personnes d'une nationalité autre que celles des pays de la communauté européenne, et elle vise à protéger aussi bien les lieux que les informations sensibles, au titre desquelles les systèmes d'information sont évidemment devenus aujourd'hui d'une importance cruciale. Ont été identifiés des *secteurs protégés*, des *unités protégées* et des unités au sein desquelles une ou des *zones à régime restrictif* (ZRR) ont été définies, incluant éventuellement des *locaux sensibles.*

L'IDRIS, concerné bien sûr directement par ces changements, est en train d'adapter ses dispositifs à la nouvelle réglementation et en premier lieu les modalités d'accès physique à la ZRR que nous avons définie, qui couvre la quasi-totalité de nos bâtiments sauf quelques salles qui seront en zone publique. Toutes les personnes qui devront accéder sans accompagnement à la zone protégée devront avoir été explicitement autorisées, soit personnellement après l'agrément de la demande qu'ils auront soumise, soit par contrat de confiance établi par les services gouvernementaux avec les sociétés prestataires qui les emploient. Les simples visiteurs placés sous la responsabilité de la personne qui les accueillera seront eux seulement soumis à une procédure déclarative. D'autre part, les changements à introduire vont concerner les accès *virtuels*, donc ceux opérés par le biais des réseaux informatiques. Dans ce cadre, nous devrons faire évoluer notre politique de sécurité selon les recommandations remises à jour de l'*Agence nationale de la sécurité des systèmes d'information* (ANSSI). Une des conséquences en sera qu'à terme tous nos utilisateurs devront être explicitement autorisés à accéder à nos services suivant la même procédure que celle décrite pour les accès physiques à notre ZRR, via donc le dépôt d'une demande explicite dont la durée de traitement ira d'une semaine pour tous les cas simples jusqu'à deux mois maximum dans certains cas complexes. Parallèlement, nous renforcerons notre vigilance sur le strict respect des engagements de sécurité que nous demandons à nos utilisateurs de prendre, en particulier ceux qui concernent l'usage impérativement personnel des informations de connexion. Nous accentuerons également nos actions pédagogiques dans le domaine de la sécurité pour convaincre chacun qu'au faible prix d'une attention maintenue en éveil et de quelques contraintes d'ampleur limitée, c'est aussi, en même temps que certains des intérêts vitaux de la nation, le fruit de leur travail qui sera mieux protégé, sans restreindre en rien les coopérations scientifiques elles-mêmes, plus que jamais vitales aujourd'hui pour l'activité scientifique et le progrès des connaissances. C'est l'objectif fécond du changement en cours de la législation, qui ne saurait être efficace et bénéfique sans l'adhésion de toutes les parties prenantes.

Ordiphones : quels enjeux pour la sécurité ?

Vincent Ribaillier - IDRIS

Aujourd'hui, de plus en plus de téléphones permettent des usages jusqu'alors réservés aux ordinateurs, en autorisant non seulement l'accès à l'Internet public mais aussi à sa messagerie personnelle et professionnelle, à n'importe quel document bureautique et plus généralement au système d'information de l'entreprise ou de l'organisme public pour lequel on travaille. Quels sont les enjeux en termes de sécurité et quels défis sont à relever par les organisations pour prévenir les risques nouveaux induits par la diffusion exponentielle de ces matériels ? Nous allons nous efforcer dans cet article d'en présenter le contexte et d'en donner des éléments de réponse.

Lorsque la société IBM dévoila en 1992 Simon, son prototype de téléphone intelligent ou *« smartphone »*, ce fut là le départ d'une longue course technologique effrénée. En effet, en intégrant dans un téléphone mobile des fonctions telles que la gestion de calendrier et de carnet d'adresses, l'envoi et la réception de fax et de courriels, l'édition de notes personnelles, IBM positionna la barre très haut vis-à-vis de ses concurrents. Cette firme prit notamment une avance technologique importante sur Apple qui présenta cette même année Newton MessagePad, le premier assistant personnel à tablette tactile mais qui n'était pas doté de fonctionnalités de téléphonie. Une nouvelle série d'offensives technologiques eut lieu en 2001 avec l'arrivée sur la scène des sociétés Palm et Microsoft, puis de RIM en 2002, qui développèrent chacune leur propre système d'exploitation pour ces équipements : Palm OS, Windows CE Pocket PC OS et Blackberry. Mais cet écosystème fut chamboulé avec l'arrivée dans la course de deux mastodontes : Apple en 2007, puis Google en 2008, qui se lancèrent dans la compétition en introduisant leurs propres systèmes d'exploitation iOS et Android.

Les *smartphones* intègrent désormais toutes les fonctions traditionnellement attribuées aux ordinateurs personnels. C'est d'ailleurs pour cette raison qu'on les désigne fréquemment sous le terme d'ordiphones. Nous assistons ainsi à une véritable révolution technologique car ces composants offrent un rapport fonctionnalités / performances / volume exceptionnel. Pesant souvent moins de 100 grammes, ils sont munis aujourd'hui de processeurs multi-cœurs cadencés à des fréquences supérieures au GHz et embarquent une multitude de capteurs (micros, caméras, puces GPS, accéléromètres) et de réseaux (GSM, Bluetooth, Wi-Fi). Cette révolution technologique ne concerne pas seulement le matériel mais également le logiciel. Les utilisateurs ont la possibilité d'accéder à des hyper-marchés d'applications, connus également sous le nom de dépôts (*« stores »*), dans lesquels plusieurs centaines de milliers d'applications téléchargeables sont disponibles, dont chacune peut être installée en un simple clic.

Dans le monde de l'entreprise, la demande pour que les ordiphones soient intégrés au système d'information (SI) est de plus en plus forte. Si ces composants apportent incontestablement un gain de productivité pour les utilisateurs, leur intégration dans le SI nécessite néanmoins d'être effectuée avec la prudence la plus extrême. En effet, du fait de la multitude des modes de connexion possibles, l'introduction de ces composants dans le SI démultiplie la surface d'attaque de ce dernier. Les nouvelles conventions d'administration et d'utilisation des ordiphones introduites par les acteurs de cet écosystème (constructeurs, développeurs, opérateurs) posent également de très sérieux problèmes et ne se prêtent pas toujours aux exigences des entreprises. Dans la plupart des cas, le principe de séparation du rôle d'administrateur et d'utilisateur n'est pas respecté si bien que l'utilisateur a souvent des privilèges d'administrateur sur l'ordiphone. De surcroît, certaines fonctionnalités souvent critiques pour la sécurité (comme la mise à jour automatique des correctifs de sécurité) sont parfois bridées par ces acteurs. Leur objectif est de rester à tout prix dans la compétition en développant le maximum d'innovations technologiques et on ne peut que constater que la sécurité est bien trop souvent négligée (absence de correctifs de mises à jour critiques dans des délais convenables, mises à jour promises par les constructeurs mais jamais diffusées dans les faits…).

La sécurité a pourtant été sérieusement pensée par les constructeurs et les développeurs de systèmes d'exploitation qui ont su l'intégrer pleinement dans leurs cahiers des charges et définir des modèles innovants aussi bien d'un point de vue matériel que logiciel. L'introduction de la mémoire Flash complexifie l'extraction des données en mémoire car, contrairement au cas des disques durs, de solides compétences en électronique sont requises ainsi que du matériel spécialisé. Du point de vue de l'architecture logicielle, les noyaux robustes du monde Unix et les concepts de virtualisation ont su être judicieusement réutilisés pour mettre en œuvre le principe du cloisonnement des applications. Ces précautions s'accompagnent d 'autres mesures de sécurité comme la gestion fine des privilèges accordés aux applications ou la signature numérique des applications installées.

Mais toutes ces précautions ne valent que si elles sont comprises de l'utilisateur final et, comme bien souvent en sécurité informatique, c'est justement là que le bât blesse ! Prenons l'exemple d'un éditeur de texte qu'un utilisateur installerait sans vérifier au préalable un à un les privilèges qu'il délègue à l'application au moment de l'installation. Cet utilisateur pourrait sans même s'en apercevoir accorder des privilèges d'accès aux données de contact ou aux différents capteurs comme la puce GPS ou le microphone, privilèges qui n'ont absolument aucune raison d'être associés à un éditeur de texte et qui par leur présence même dans la liste des privilèges à attribuer rendent l'application hautement suspecte. De la même manière, la confiance qu'il faut accorder à un dépôt d'applications grand public réputé sûr ne doit être que très limitée. En effet, comment techniquement les applications déposées sur un dépôt peuvent-elles être sérieusement auditées par les gestionnaires de celui-ci lorsque la fréquence journalière de soumission d'applications est de l'ordre du millier ?

Le catalogue des menaces recensées est très vaste et des concepts très plausibles d'attaques à grande échelle redoutables (visant par exemple les dépôts d'applications) ont été décrits par les chercheurs en sécurité informatique. Parmi les menaces les plus fréquemment observées, citons les attaques exploitant la faiblesse ou l'absence des mots de passe ou motifs de déverrouillage des sessions utilisateurs, les attaques d'atteintes à la confidentialité des données (fuites de données suite à une perte ou un vol de l'ordiphone ou suite à l'installation d'un logiciel malveillant ou même tout simplement suite à un rechargement de l'ordiphone sur une prise USB malicieuse, comme certains cas en ont déjà été rapportés), les techniques d'usurpation d'identité par la mise en place d'une antenne GSM malicieuse, l'espionnage de l'utilisateur suite à l'installation d'une application autorisée à accéder aux différents capteurs embarqués (microphone, caméra, GPS). N'oublions pas non plus les menaces accidentelles ou délibérées d'atteinte à la disponibilité du réseau GSM. En octobre 2011, lorsque le réseau RIM a été défaillant pendant plusieurs jours, les millions d'utilisateurs d'ordiphones Blackberry ont pu par exemple mesurer le désagrément que ce type d'incident peut occasionner. Enfin, les opérations de débridage (*« jail breaking »*) parfois suggérées aux utilisateurs pour contourner certaines restrictions ne sont pas sans risque car elles peuvent court-circuiter des briques essentielles de sécurité.

Face à ce constat, la tâche n'est donc pas aisée pour l'entreprise ou le laboratoire désirant intégrer ces composants dans son SI. L'IDRIS a eu récemment l'occasion de réfléchir à cette problématique, d'une part pour déployer un parc de taille limité d'ordiphones dans son SI et d'autre part dans un cadre plus général de réflexion avec la *Coordination régionale de la sécurité des systèmes d'information* (CRSSI) de la *Délégation régionale d'Île-de-France Sud* du CNRS. Ce travail a permis d'élaborer des recommandations simples qui sont résumées de façon très synthétique dans le paragraphe qui suit.

Tout d'abord, il est recommandé d'intégrer les ordiphones dans le périmètre des actifs d'information entrant dans l'analyse de risques SSI. Cette démarche permet de recenser les différentes menaces puis de fixer les objectifs de sécurité en fonction du contexte, c'est-à-dire de définir les risques qui peuvent être acceptés ainsi que ceux qui doivent être évités. La prémunition de la fuite d'informations considérées comme sensibles pourra par exemple constituer un impératif de sécurité alors qu'au contraire la perte de disponibilité d'une application pourra être considérée comme un risque acceptable. Une fois les objectifs de sécurité établis, il convient d'identifier les mesures de sécurité permettant de les atteindre. Les mesures de sécurité traditionnelles basées sur la mise en place de dispositifs techniques ne sont malheureusement pas toujours faciles à identifier et à déployer dans l'univers des ordiphones. Il est possible dans ce cas de recourir à la mise en place de mesures de sécurité de nature réglementaire destinées à responsabiliser les utilisateurs. La configuration du chiffrement des données est par exemple dans certains cas tellement complexe à mettre en œuvre qu'il peut être préférable d'interdire dans la politique de sécurité du système d'information (PSSI) l'enregistrement de données sensibles sur les ordiphones. Dans tous les cas, la mise en place des mesures de sécurité devra être envisagée avec l'accompagnement d'une démarche de sensibilisation des utilisateurs. Si l'offre des solutions de sécurisation d'ordiphones n'est d'une façon générale pas encore très mature pour une utilisation en environnement professionnel, elle est en revanche en pleine effervescence et il y a fort à parier que des solutions innovantes verront prochainement le jour. Les premières solutions anti-virales pour ordiphones viennent par exemple de faire leur apparition. Les organisations souhaitant déployer un parc d'ordiphones de taille importante auront tout intérêt à comparer les différentes solutions professionnelles de gestion de parc de composants mobiles dites MDB (*« Mobile Device Management »*) qui, moyennant un coût d'entrée certes non négligeable, apportent une vraie plus-value en termes de sécurité. Les lecteurs pourront se référer aux documents suivants pour une analyse plus détaillée des mesures de sécurité qu'il est possible de mettre en œuvre.

Prototype Simon présenté en 1992

On voit donc que, face au déploiement rapide de ces merveilles technologiques que sont les ordiphones, certes à la source de nouveaux usages très attractifs, offerts qui plus est dans un grand confort d'utilisation, des enjeux majeurs de sécurité sont posés à tous les organismes professionnels. Le CNRS en est bien sûr pleinement conscient et réfléchit à cette problématique, notamment à travers ses instances nationales et régionales dédiées à la sécurité des systèmes d'information, instances auxquelles l'IDRIS joint, dans ce domaine-là comme dans d'autres, son expertise.

Références

Article Wikipédia sur les *smartphones*, http://en.wikipedia.org/wiki/SmartphoneArticle Wikipédia sur le PDA Newton, http://fr.wikipedia.org/wiki/Newton_PDA
Smartphones: Information security risks, opportunities and recommendations for users, ENISA, December 2010, http://www.lsec.be/upload_directories/documents/ENISA/ENISA_Smartphone_Security.pdf
Appstore security, 5 lines of defense against malware, ENISA, September 2011, http://www.enisa.europa.eu/act/application-security/smartphone-security-1/appstore-security-5-lines-of-defence-against-malware
MISC N°51 (septembre/octobre 2010), dossier « Sécurité des OS mobiles »
MISC N°57 (septembre/octobre 2011), article « Introduction au reverse engineering d'application iOS : déchiffrement et analyse statique d'ARM »
MISC N°58 (novembre/décembre 2011), article « Renforcez la confidentialité de votre Android »
MISC N°58 (novembre/décembre 2011), article « Analyse d'une des nouvelles menaces pour Android : Trojan.AndroidOS. Dogowar.a »
Exemple de scénario de compromission d'un iPhone via un dépôt d'applications, http://www.youtube.com/watch?feature=player_embedded&v=ynTtuwQYNmk
The Most Vulnerable Smartphones of 2011, BIT9, 21 novembre 2011, http://www.bit9.com/files/Bit9Report_SmartPhones2011.pdf

Extension de la capacité des infrastructures techniques de l'IDRIS

Rafael Medeiros - IDRIS

L'IDRIS a procédé en 2011 à une augmentation très significative de la capacité de ses infrastructures techniques, en multipliant celle-ci par un facteur 2,5.

Après une première phase de travaux effectuée en 2008, qui avait alors permis de rénover et adapter l'infrastructure aux besoins des configurations informatiques installées cette année-là et actuellement en place, les travaux de 2011 ont permis de préparer notre centre pour l'arrivée de la prochaine génération de nos supercalculateurs.

Ces travaux, réalisés en site occupé, nous ont permis de fonctionner normalement pendant toute leur durée, hormis les deux arrêts techniques du printemps et de l'été derniers, indispensables pour permettre de basculer la source d'énergie de l'ancien réseau du campus de l'université vers notre nouveau réseau haute tension dédié.

Les faits majeurs relatifs à ces travaux ont été :

— la création d'un poste de livraison haute tension, nécessitant la mise en place d'un câble haute tension dédié, d'une longueur de 5 km ;
— la création d'un nouveau poste de transformation électrique et la consolidation du poste existant ;
— l'installation d'une nouvelle ligne d'onduleurs et la consolidation de la ligne déjà en exploitation ;
— l'installation de trois nouveaux appareils de production d'eau glacée, en complément des trois déjà opérationnels ;
— la mise en place d'un programme de délestage de la salle machines, permettant aux groupes électrogènes, dont la capacité globale n'a pas été augmentée et qui ne pourront donc plus couvrir l'ensemble de nos besoins électriques, de secourir les équipements vitaux du centre en cas de coupure totale du réseau haute tension ;
— la mise en place d'un système de récupération de chaleur permettant d'alimenter, dès 2012, les réseaux de chauffage de nos bâtiments et plus tard ceux d'autres bâtiments CNRS voisins.

Les chiffres clés des travaux et de notre infrastructure d'aujourd'hui sont :

— 3 ans : durée totale de cette opération, depuis l'étude de faisabilité jusqu'à la réception des équipements ;
— 11 mois : durée des travaux proprement dits pour l'installation des différents équipements et leur mise en service ;
— 8 entreprises (hormis leurs sous-traitants) : soit plusieurs dizaines de personnes impliquées dans ces travaux ;
— 2,5 MW : puissance redondante secourue par des onduleurs disponible en salle machines pour l'alimentation électrique des calculateurs et des autres serveurs ;
— environ 4 MW : puissance électrique totale qui sera utilisée par notre centre de calcul dans les prochaines années, pour alimenter à la fois les calculateurs installés en salle machines et l'ensemble des équipements de l'infrastructure technique.

Groupes de production d'eau glacée

Réseau de distribution d'eau glacée

Onduleurs pour la protection des salles machines

Groupe électrogène pour la protection des salles machines

Informations

Actualité

Sélection des *PRACE Advanced Training Centers*

PRACE vient de sélectionner six organisations en Europe pour mettre en œuvre des centres de formation dédiés au calcul de haute performance (les *PRACE Advanced Training Centers*). L'IDRIS, aux côtés de ses partenaires du CEA-CCRT/TGCC, du CINES et de l'INRIA, assurera, sous la coordination de la Maison de la simulation, la participation française.
Pour en savoir plus : http://www.prace-ri.eu/PRACE-to-establish-six-advanced

Calendrier des formations IDRIS programmées d'ici l'été 2012

Titre de la formation	Date de début	Durée
Calcul parallèle : OpenMP	13/03/2012	2 jours
Calcul parallèle : MPI	21/05/2012	4 jours
Programmation hybride MPI/OpenMP	02/04/2012	4 jours
Fortran de base : Fortran 95-1	03/04/2012	3 jours
Fortran : Fortran 95-2	20/03/2012 et 05/06/2012	3,5 jours
Fortran 2003	19/06/2012	3 jours
Langage C	25/06/2012	5 jours

Ces dates vous sont communiquées à titre d'information et sont susceptibles d'être mises à jour. Pour une information récente et plus complète, n'hésitez pas à consulter le serveur Web des cours de l'IDRIS : https://cours.idris.fr
Vous y trouverez le catalogue complet des formations et pourrez vous pré-inscrire aux sessions annoncées.

Nous vous rappelons que les formations IDRIS sont gratuites pour les personnes appartenant au CNRS ou à une université. Elles sont aussi accessibles au personnel d'entreprises publiques ou privées via CNRSFormation Entreprises : les conditions d'inscription sont alors consultables sur le site Web : http://cnrsformation.cnrs-gif.fr

Directeur de la publication : Denis Girou
Rédacteur en chef : Thierry Goldmann
Rédactrice adjointe : Geneviève Morvan
Comité de rédaction : Denis Girou, Thierry Goldmann, Pierre-François Lavallée, Geneviève Morvan
Conception graphique, réalisation et impression : Graficom – tél 01 69 51 02 99 – www.graficom.pro

IDRIS – Institut du développement et des ressources en informatique scientifique
Rue John von Neumann
Bâtiment 506 – BP 167
91403 ORSAY Cedex – tél +33 (0)1 69 35 85 00 – www.idris.fr