Contributions à l'optimisation combinatoire pour l'embarqué : des autocommutateurs cellulaires aux microprocesseurs massivement parallèles

Soutenance de thèse d'Habilitation à Diriger des Recherches

Renaud Sirdey (CEA)

Institut de Physique Théorique, 29 novembre 2011

Parcours

- 1998 : Diplôme d'ingénieur de l'UTC.
- 1998 : MSc en math. appli. de l'Université de Cranfield (major), G.-B.
  - Thèse de MSc (1998 Texas Inst. MSc Thesis Prize for DSIP).
- 09/1998 : Member of Scientific Staff, Nortel PND, Maidenhead, G.-B.
- 07/2001 : Responsable d'équipe R&D, Nortel GSM Access R&D, Châteaufort.
- 09/2003 : Architecte système, Nortel GSM Access R&D, Châteaufort.
  - 03/2007 : Thèse de doctorat (prix Guy Deniélou 2008).
    - Directeurs : J. Carlier, D. Nace (UTC).
    - Rapporteurs : P. Baptiste (Polytechnique), A. R. Mahjoub (Dauphine).
- 11/2007 : Ingénieur-chercheur, CEA, Saclay.
  - 01/2010 : Nommé expert du CEA.
  - 11/2011 : Soutenance de thèse d'HDR.

Publications

- 9 articles dans des journaux internationaux.
  - Eur. J. Oper. Res. (1), J. Heuristics (1), Comput. Oper. Res. (1), 4OR Q. J. Oper. Res. (3), RAIRO Oper. Res. (1), Int. J. Innov. Comput. Appl. (1), Icarus (1).
- 5 articles dans des conférences internationales.
  - ICA3PP'11, COCOON'11, RTSS'10, ISCO'10, INOC'03.
- 5 brevets.
  - 2 Nortel (délivrés), 3 CEA (déposés).
- 6 articles de vulgarisation.
  - Dont 3 articles dans le magazine La Recherche.
- 2 workshops internationaux.
  - ADPNA'11, ICE'10.
- Divers communications, rapports techniques et séminaires.

Activités d'encadrement

- Architecte système (période 2003-07).
  - Co-responsable scientifique et technique des activités de R&D pour un autocomm. réparti de nouvelle génération.
  - Principaux domaines d'activités : algorithmique, algorithmique répartie, optimisation discrète, fiabilité des systèmes, statistiques, génie logiciel, cryptologie et sécurité, traitement du signal.
- Responsable d'équipe de recherche (période 2008-)
  - Responsable de l'équipe "Calcul intensif embarqué" du LaSTRE représentant une dizaine de chercheurs spécialisés en compilation et en logiciel système pour les architectures parallèles.
  - Principaux domaines d'activités : parallélisme, compilation, optimisation combinatoire, preuve de programmes, traitement du signal et de l'image, cryptologie.
- Co-encadrement de doctorants.
  - 2008-11 : thèse de Sergiu Carpov (CEA/UTC) en ordonnancement.
    - Soutenue le 14 oct. 2011, Sergiu est maintenant chercheur au CEA.
  - 2010-13 : thèse de Oana Stan (CEA/UTC) en optimisation sous incertitude.
  - 2011-14 : thèse de Simon Fau (CEA/UBS) en cryptologie.
- Encadrement de plusieurs stages ingénieur ou master.

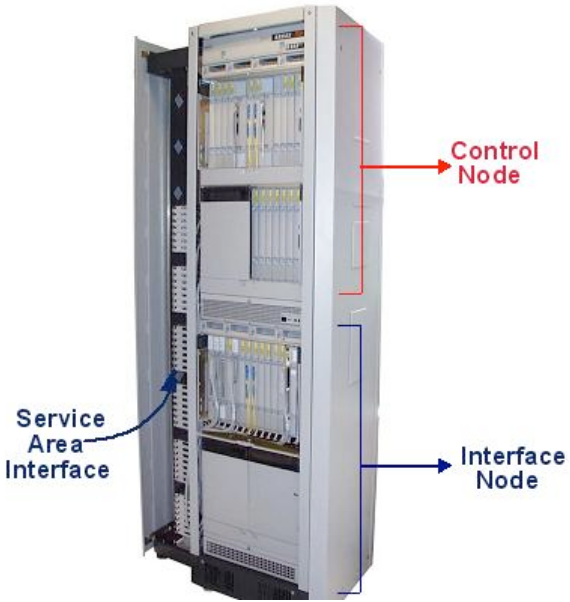
Enseignement & expertise

- Activités d'enseignement :
  - 2007- : professeur chargé de cours à l'ENSTA, responsable du module IN204.
  - 2005- : chargé d'enseignement vacataire à l'UTC, module RO03.
  - Interventions occasionnelles sur les applications industrielles de la RO (X, Panthéon-Sorbonne, ...).
- Autres activités :
  - Coordinateur du thème transversal "Recherche opérationnelle" et responsable du séminaire LaSTRE.
  - Membre extérieur du conseil de l'école doctorale de l'UTC.
  - Membre du réseau d'experts CEA-Technologie-Conseil.
  - Expertises ponctuelles (interne CEA, OSEO, ANR, ...).
  - Rapporteur occasionnel pour les journaux Eur. J. Oper. Res., Discrete Appl. Math., Discrete Optim. et pour diverses conférences (récemment DRCN 2007, INOC 2009, EMSOFT 2009, ISCO 2010, ...).

Optimisation & autocommutateurs (2001-07)

Qu'est-ce qu'un autocommutateur ?

- Équipement en charge de l'écoulement du trafic dans un réseau cellulaire.
- Un autocommutateur typique :
  - Plusieurs dizaines de modules.
  - Plusieurs millions de lignes de code.
  - Une dizaine d'interfaces.
  - Des mécanismes avancés pour la tolérance aux pannes et la maintenance.
    - Différents modes de redondance.
    - Montée en charge "plug & play".
- C.-à-d., un système complexe offrant des ressources interdépendantes qu'il convient d'utiliser efficacement.



Le BSCe3, un BSC GSM haute capacité.

Problèmatiques combinatoires

- Affectation de canaux radio à des trans-récepteurs.
  - Flots avec combinaisons lexicographiques de critères.
- Affectation de cellules radio à des liens PCM avec demandes excédentaires en circuits.
  - Bin-packing extensible et... Répartition de sièges.
- Maximisation de la durée de vie sur batterie d'une BTS multicabinet.
  - Sac à dos max-min multidimensionnel.
- Affectation de tâches sous contraintes de tolérance aux pannes.
  - k-way partitioning.
- Configuration dynamique de DSP.
  - Flots avec combinaisons lexicographiques de critères (calcul de QoS).
- Ordonnancement de migrations de tâches (*thèse de doctorat*).
  - Ordonnancement sous contraintes de ressource.
- **Synchronisation d'horloges.**

Synchronisation d'horloges

- Problématique liée à l'introduction de technologies de transport par paquets (bas coût) dans les réseaux GSM (entre les BSC et les BTS).
- Difficulté : pour qu'il fonctionne correctement (handovers en particulier) les BTS d'un réseau GSM doivent disposer d'horloges très précisément synchronisées ($\pm$50 PPB).
  - Service garanti sur un backhaul TDM.
  - Aucune garantie sur un réseau paquet asynchrone.
- Solutions envisagées :
  - Récepteur GPS à la BTS.
  - Horloge au rubidium à la BTS.
  - Véhiculer des données de synchronisation au travers des flux de paquets temps réel (appels voix paquétisés).

État de l'art

- Mesures monodirectionnelles :
  - Estimation par régression linéaire avec intervalles inter-émissions en abscisse et inter-réception en ordonnée (peu précise).
  - Estimation par positionnement d'une droite sous un nuage de points de mesure avec délais d'acheminement en abscisse et dates d'envois en ordonnée [Moon et al. ; Zhang et al.]
- Mesure bidirectionnelles :
  - Duda et al. : robustesse inhérente à l'utilisation de mesures bidirectionnelles mais hypothèse (peu réaliste) d'un délai d'acheminement minimum négligeable.
- Autres approches :
  - Aweya et al. (paquets de taille fixe sans données additionnelles), Khlifi et Grégoire (faible exigence de précision), ...

Principes généraux

- Problème ouvert de l'état de l'art (académique et industriel) : respecter l'exigence de précision élevée (mais normative) du GSM.
  - Stratégie systémique : mesures bidirectionnelles, datation haute précision, méthode d'estimation robuste.
- Dater les paquets à l'émission et à la réception au BSC et à la BTS.
  - À l'aide d'un timestamper matériel, au plus proche de l'interface réseau.
  - Paquet de follow-up (à la IEEE 1588).
  - Redescente des paires de dates du flux BTS vers BSC à la BTS.
- Particularité : on ne s'intéresse qu'au biais d'horloge c.-à-d. au coefficient directeur de l'équation de droite liant les deux horloges.
- Accumuler des données (1 paquets toutes les 5 ms) pour une estimation robuste—à la BTS—du biais par rapport à l'horloge du BSC.
  - Utiliser l'estimation pour appliquer une correction (bornée) à l'oscillateur local de la BTS.

Estimation par programmation linéaire

– Lien entre les deux horloges,

$$(1) \quad t^{(\mathrm{S})} = \alpha t^{(\mathrm{M})} + \beta,$$

où β est le décalage, $\alpha \approx 1$ est le biais, et où la dérive est négligée (phén. long terme).

– Pour un paquet descendant on a

$$t_i^{(\mathrm{S})} = \alpha t_i^{(\mathrm{M})} + \beta + \alpha D_i^{(d)},$$

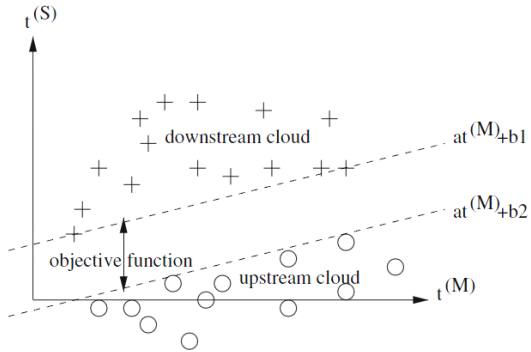
i.e., un point au dessus de (1).

– Pour un paquet montant on a

$$\tau_j^{(\mathrm{S})} = \alpha \tau_j^{(\mathrm{M})} + \beta - \alpha D_j^{(u)},$$

i.e., un point en dessous de (1).

– Estimation du biais par séparation des deux nuages de points à l'aide d'un "couloir" aussi large que possible.



$$\begin{cases} \text{Maximiser } \beta_1 - \beta_2 \\ \text{s. l. c.} \\ \alpha t_i^{(\mathrm{M})} + \beta_1 \leq t_i^{(\mathrm{S})}, \ i \in \{1, \ldots, n\}, \\ \alpha \tau_j^{(\mathrm{M})} + \beta_2 \geq \tau_j^{(\mathrm{S})}, \ j \in \{1, \ldots, m\}. \end{cases}$$

Validation expérimentale

- Une durée d'accumulation des données d'une dizaine de minutes permet une estimation du biais avec une précision de quelques PPB.
  - Y compris dans des conditions réseaux défavorables.
- Robustesse de la méthode :
  - Utilisation de mesures bidirectionnelles.
  - Et robustesse de l'estimation par programmation linéaire.
- Processus de validation expérimentale et industrielle :
  1. Simulations [4OR Q. J. Oper. Res. 6(4), 2008].
  2. En laboratoire avec simulateurs de trafic.
  3. En prédéploiement puis en opérations sur le terrain.
- Brevet WO/2010/083930 sur la méthode dans sa globalité.
- Méthode ultérieurement adaptée par Poirier et al. pour la synchronisation hors ligne des traces d'exécution d'un système réparti.

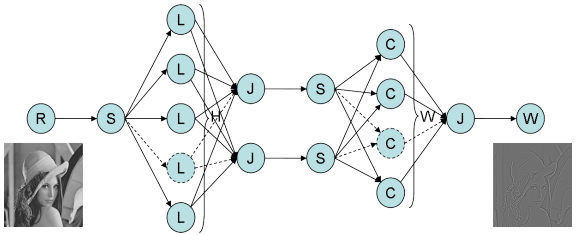
Optimisation & compilation (2007-)

Compilation & recherche opérationnelle

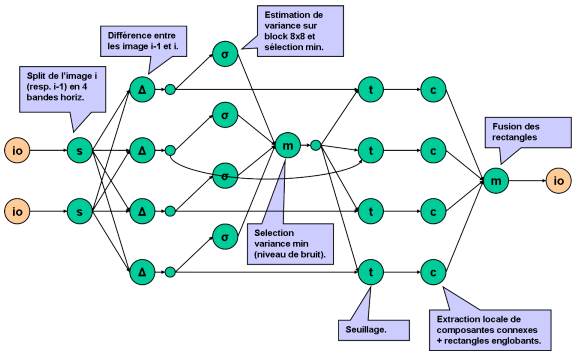
- Depuis toujours, la RO a des applications en compilation, principalement au niveau de la phase de production du code machine, le backend.
  - Notamment pour l'exploitation efficace du parallélisme d'instructions.
- Fin de l'ère Moore oblige, la nouvelle génération de microprocesseurs n'a rien à envier aux supercalculateurs des années 90 !
  - Ensemble de calculateurs parallèles symétriques à mémoire partagée interconnectés par un réseau paquet asynchrone... Sur puce.
  - D'où une complexification du processus de compilation.
    - En particulier, plusieurs problèmes délicats d'affectation de ressources se dressent entre un programme et son exécution performante et maîtrisée.
- Dès lors, la pertinence des techniques de RO se généralise bien au-delà du seul backend.
  - Difficultés : taille des instances, incertitudes, multicritère.

Programmation flot de données

- Modèles flot de données.
  - Réseaux de tâches communiquant par des canaux FIFO.
  - Synchronisation par les données.
- Le langage ΣC.
  - Adapté à une large gamme d'applications embarquées.
  - Approche par composants pour la maîtrise d'applications complexes.
  - Expression naturelle et explicite du parallélisme.
  - Programmation proche de la machine possible si nécessaire.
  - Extension du langage C.



Réseau de calcul du laplacien d'une image.

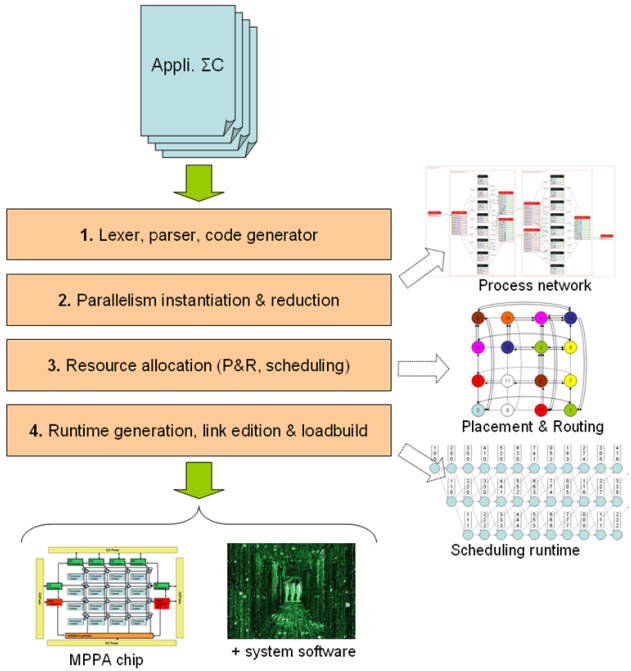


Appli. de détection et suivi de cibles.

LNCS 7016 (ICA3PP'11), 2011.

Processus de compilation

- Analyses lexicales, syntaxique, etc.
- Génération des codes d'instanciation et de traitement.
- Construction du réseau de processus.
- Composition des schémas d'accès aux données, réduction de parallélisme.
- Dimensionnement des tampons.
- Calcul d'un ordre partiel d'exécution.
- **Placement/routage.**
- Génération des données de runtime.
- Construction du binaire.
- Exécution et... Compilation itérative.



Compilation d'une appli. ΣC.

Problématique du placement/routage

- Partitionnement :
  - Grouper, sous des contraintes de capacités multidimensionnelles, les tâches qui ont tendance à communiquer ensemble.
  - Optimisation d'un critère de débit réseau.
- Placement :
  - Affecter les groupes de tâches qui communiquent à des éléments d'architecture proches (au sens d'une distance de routage).
  - Optimisation d'un critère de latence.
- Routage :
  - Calculer les chemins d'acheminement des données au travers du réseau sur puce.
  - Optimisation d'un critère de latence et de charge des liens.
- Problèmes sous-jacents NP-difficiles.

Approches du problème

- Partionnement puis placement puis routage.
  - Décomposition qui facilite la résolution du problème mais déstructure.
    - Conflits d'objectifs.
    - Problèmes éventuels de réalisabilité en aval.
- Première approche pragmatique.
  - Décomposer et résoudre chaque problème indépendamment à l'aide d'algorithmes approchées et rapides (ou résolution exacte si possible).
    - Approche adaptée au début du cycle de développement.
    - Conception d'algorithmes qui pourront être utilisés dans une méthode globale.
- Approche globale.
  - Problème plus complexe, intrinsèquement multi-critère.
  - Augmentation significative des temps calcul, approche inadaptée au début du cycle de développement.

Partitionnement

– Soit $G = (V, A)$ un graphe orienté, soit R un ensemble de ressources et N un ensemble de nœuds. Sont également donnés une fonction de taille $s : V \longrightarrow \mathbb{R}^{+|R|}$, une fonction d'affinité $q : A \longrightarrow \mathbb{R}^{+}$ et un vecteur de capacité nœud $C \in \mathbb{R}^{+|R|}$.

– Problème : trouver une affectation (totale) des sommets aux nœuds, $f : V \longrightarrow N$, qui respecte les contraintes de capacité

$$\sum_{v \in V : f(v) = n} s(v) \leq C, \forall n \in N,$$

et qui minimise la fonction économique

$$\sum_{a=(v,w) \in A : f(v) \neq f(w)} q(a).$$

– Complexité : NP-difficile au sens fort.

État de l'art

- Résolution approchée.
  - Kernighan & Lin : bipartitionnement, généralisation pseudopolynomiale au cas pondéré.
  - Recuit simulé [Johnson et al.] : fonction économique à termes de pénalité.
  - Algo. glouton par affinité relative et terme "gravitationnel" [David et al.].
- Résolution exacte.
  - Approche polyèdrale [Ferreira et al.].
    - Résolution possible jusqu'à de l'ordre de 300 sommets et 500 arcs.
- Bornes.
  - Relaxations linéaire et semi-définie [Lisser & Rendl].
    - Mettent en jeu $O(n^2)$ variables et $O(n^3)$ contraintes.
- Taille des instances : jusqu'à de l'ordre de 4000 sommets et d'une soixantaine de partitions.

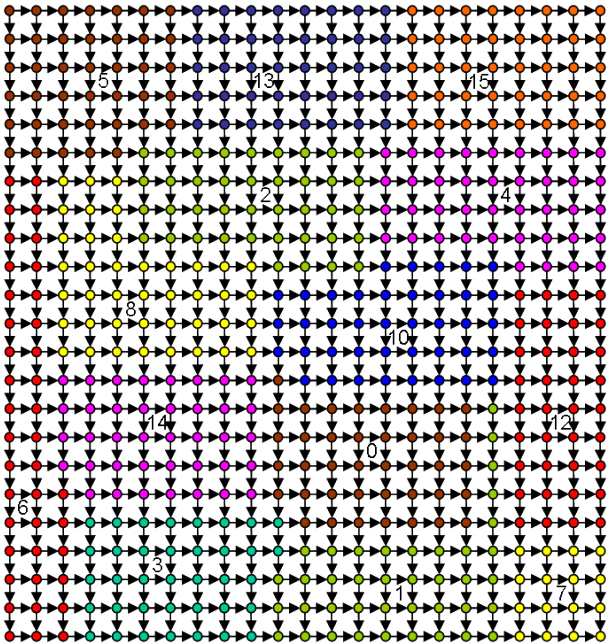
Résolution par demi-GRASP

– Affinité relative de $S \subset T$ envers $S' \subset T$ $(S \cap S' = \emptyset)$:

$$\propto \alpha(S, S') \left(\frac{1}{\alpha(S, \bar{S})} + \frac{1}{\alpha(S', \bar{S}')} \right),$$

où $\alpha(S, S') = \sum_{a \in \delta(S,S')} q_a$.

– Algorithme par construction progressive :
  – Choix de l'affectation (admissible) de tâche ou de la fusion (admissible) de partition d'aff. rel. maximale.
  – Diversification par randomisation et démarrages multiples.



Partitionnement d'une grille 23×23.

Partitionnement stochastique

- Nécessité de prendre en compte le caractère incertain des poids des tâches (temps d'exécution).
  - Lois jointes complexes, intrinsèquement multimodales.
  - D'où un besoin de méthodes non paramétriques et d'utilisation explicite de données expérimentales.
- Principe : approche par scénarios interprétée dans le cadre de la théorie des tests d'hypothèses statistiques.
- S'intègre aisément et efficacement dans des algorithmes de résolution approchée existants.

Exemple : reformulation mixte de

$$\begin{cases} \text{Minimiser } c^T x, \\ \text{s. l. c.} \\ P(Ax \leq b) \geq 1 - \varepsilon, \end{cases}$$

en,

$$\begin{cases} \text{Minimiser } c^T x, \\ \text{s. l. c.} \\ \tilde{A}_i x \leq b + (1 - \chi_i) L, \qquad \forall i, \\ \sum_{i=1}^{N} \chi_i \geq k(N, 1 - \varepsilon, \alpha), \end{cases}$$

où α est un niveau de confiance.

Modification de l'algorithme

– Respect de la contrainte en probabilité

$$P\left(\bigwedge_{n\in N}\bigwedge_{r\in R}\sum_{v\in V:f(v)=n} S_{vr} \leq C_r\right) \geq 1-\varepsilon,$$

avec un niveau de confiance α.

– Remplacement du test d'admissibilité, par exemple pour l'affectation du sommet v à la partition n,

$$S_{vr} + \sum_{w\in V\setminus W:f(w)=n} S_{wr} \leq C_r, \forall r \in R,$$

par

$$\sum_{k=1}^{N} \chi\left(\left\{\exists n' \neq n, \exists r : \sum_{w:f(w)=n'} \tilde{S}_{wr}^{(k)} > C_r\right\} \vee \left\{\exists r : \tilde{S}_{vr}^{(k)} + \sum_{w:f(w)=n'} \tilde{S}_{wr}^{(k)} > C_r\right\}\right) \leq N - k(N, 1-\varepsilon, \alpha),$$

– Accroissement de complexité en $O(N)$.

Placement

– Soit $H = (N, B)$ le graphe réduit et soit K l'ensemble des éléments d'architecture ($|N| = |K|$). Est également donnée une matrice de distance D.
  – Plan de routage donné a priori, au moins sur le plan des distances.
– Problème : trouver une bijection $g : N \longrightarrow K$ qui minimise

$$\sum_{(n,m)\in B} D_{g(n)g(m)}.$$

– Complexité : NP-difficile au sens fort (QAP).
  – Résolution exacte possible jusqu'à de l'ordre de 30 nœuds.
– Taille des instances : 16 à 64 nœuds.
– Résolution pratique à l'aide d'un recuit simulé par approximation différentielle.
  – C.-à-d. visant (par construction) à simuler une loi stationnaire telle que

$$P(c(\omega) \leq e_1 + \beta(e_P - e_1)) \geq \alpha.$$

Calcul des chemins de routage

- Modèle de multi-flot.
- Routage par plus court chemin.
  - $\gamma_{fa} = 0$ s'il existe un plus court chemin de $s(f)$ à $d(f)$ passant par a, 1 sinon.
  - Une solution de coût 0 indique que tout le trafic est acheminé via des plus courts chemins.
  - Une solution de coût > 0 quantifie la déviation.
- Contraintes de chemins uniques.
  - Modèle mixte, NP-difficile.
  - Mais instances pratiques accessibles avec COIN-CBC.

$$
\begin{cases}
\text{Minimiser } \displaystyle\sum_{a \in A} \sum_{f \in F} \gamma_{fa} x_{fa}, \\
\text{s. l. c.} \\
\displaystyle\sum_{a \in \delta^+(s(f))} x_{fa} = w(f), \forall f \in F, \\
\displaystyle\sum_{a \in \delta^-(d(f))} x_{fa} = w(f), \forall f \in F, \\
\displaystyle\sum_{a \in \delta^-(v)} x_{fa} = \sum_{a \in \delta^+(v)} x_{fa}, \forall f, \forall a \in A \setminus \{s(f), d(f)\}, \\
\displaystyle\sum_{a \in \delta^-(v)} \chi_{fa} \leq 1, \forall f \in F, \forall v \in V, \\
\displaystyle\sum_{a \in \delta^+(v)} \chi_{fa} \leq 1 \forall f \in F, \forall v \in V, \\
0 \leq x_{fa} \leq \chi_{fa} C_a, \forall f \in F, \forall a \in A, \\
\chi_{fa} \in \{0, 1\}, \forall f \in F, \forall a \in A.
\end{cases}
$$

Résolution globale

- Principe : chercher à intégrer les algorithmes précédents dans un ensemble cohérent.
  - Problèmes maître (partitionnement) et esclave (placement/routage).
- À chaque étape de l'algorithme par construction progressive, on dispose d'un partitionnement partiel admissible sur le plan des contraintes de capacité.
  - Inclure un placement/routage (au moins sur le plan de l'admissibilité) dans le critère d'admissibilité de l'affectation de tâches et de fusion de partitions.
  - Evaluation des choix en parallèle (sans problème de granularité) et parallélisme multistart.
- Temps de calcul de quelques dizaines de minutes.
  - Acceptable, dans l'embarqué, en fin de cycle de développement.

Conclusion & perspectives

- Côté "marche aléatoire".
  - Généralisation des algorithmes pour le placement/routage au multi-puce et à la prise en compte des défaillances et de la consommation énergétique.
  - Dimensionnement de réseaux flot de données généraux [EPTCS 38, 2010].
  - Optimisation du préchargement des données [Comput. Oper. Res. 39(3), 2012].
  - Minimisation des préemptions et des migrations dans les ordonnancements temps réel préemptifs [RTSS'10].
- Programmation stochastique, optimisation robuste et... Parallélisme.
  - Thèse d'O. Stan. dont l'objectif est de revisiter une partie des problèmes d'optimisation posés en compilation sous l'angle incertain [ETR'11].
  - Et recours au parallélisme pour mitiger l'augmentation de complexité des modèles et des algorithmes (stage B. Maurin, thèse 2012).

Conclusion & perspectives

Parallélisme embarqué.

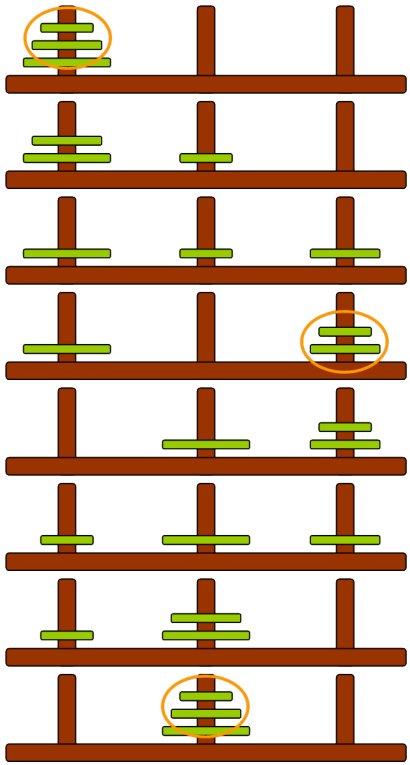
- Compilation itérative : infrastructure de remontée de données d'exécution aux phases d'affectation des ressources.
- Réseaux de tâches à topologie dynamique : reconfigurations ou... Défaillances.
- Hétérogénéité applicative : calcul intensif asservi à un noyau temps réel.

Cryptologie.

- Thèse de S. Fau dont l'objectif est d'exploiter des avancées théoriques récentes pour une mise en pratique concrète des systèmes de chiffrement algébriquement homomorphiques.
  - C.-à-d. permettant d'additioner et de multiplier des clairs en ne manipulant que des chiffrés.
- Infrastructures d'authentification et d'exécution sécurisée traitant des flux d'instructions et de données chiffrés.

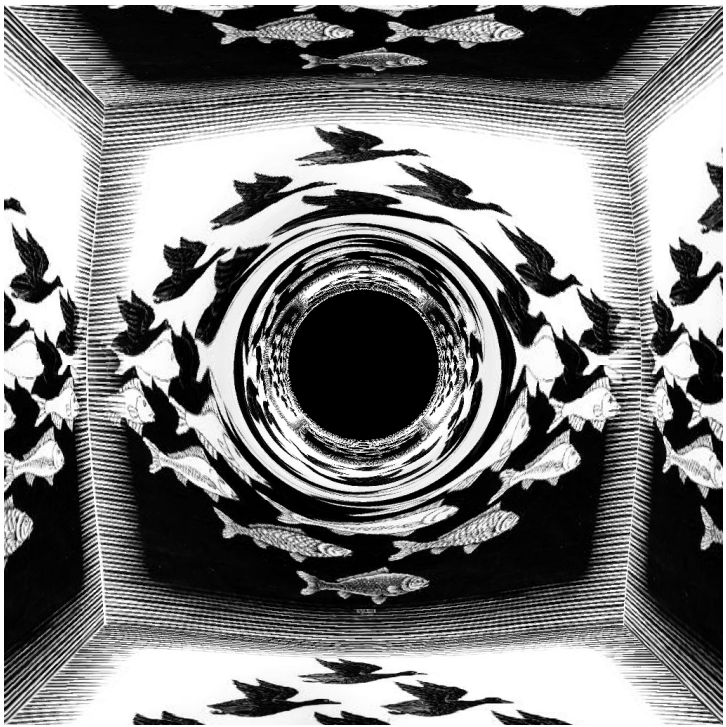
Diffusion du savoir scientifique

- Vulgarisation dans *La Recherche*.
  - Approche polyèdrale (juillet 2007).
  - Logique de Hoare (octobre 2007).
  - Consensus réparti (avril 2008).
- Conception d'ateliers pour la fête de la science à l'école primaire de Cernay-la-Ville.
  - Raisonnement par récurrence sur les tours de Hanoï.
  - Optimisation de la tournée du Père Noël en vallée de Chevreuse.
  - Dualité coloriage/clique maximum en Europe.
  - Résolution optimale en nombre de rotations du Rubik's Cube 2×2.



Clin d'œil à l'IPhT

$$\frac{dr}{dt} = \pm\left(1-\frac{2M}{r}\right)\sqrt{1-\left(1-\frac{2M}{r}\right)\frac{b^2}{r^2}},$$
$$\frac{d\phi}{dt} = \pm\frac{b}{r^2}\left(1-\frac{2M}{r}\right).$$



Merci !