

# Projet « DOLPHIN »



**D**iscrete multi-objective **O**ptimization for **L**arge scale  
**P**roblems with **H**ybrid **d**istributed tech**N**iques

**Responsable** : E-G. Talbi

**Composition** : 1 **Prof** : E-G. Talbi  
3 **MC** : C. Dhaenens, N. Melab  
F. Seynhaeve

**7 thésards en cours** : B. Weinberg, N. Jozefowicz,  
M. Basseur, S. Cahon, M. Khabzaoui, M-S. Mezmaz,  
J. Lemesre

**1 Postdoc** : L. Jourdan

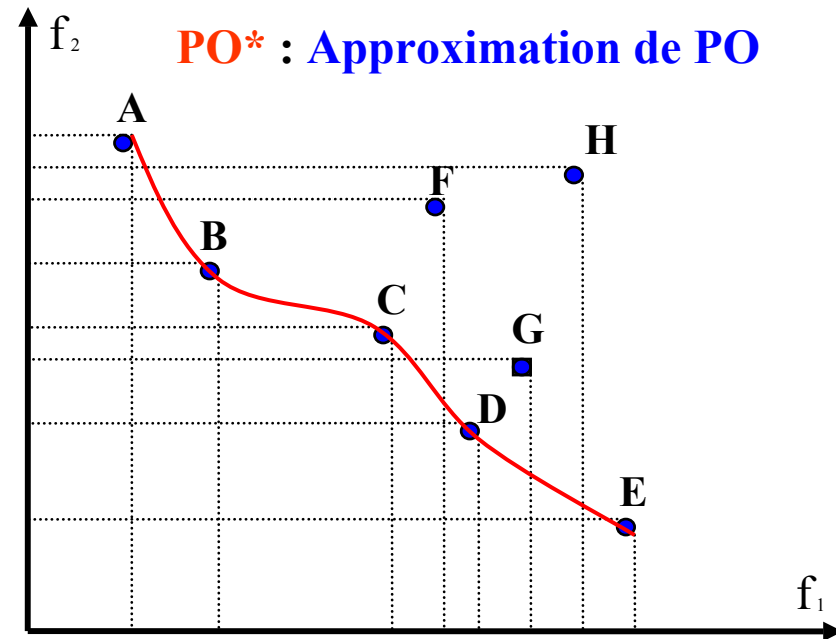
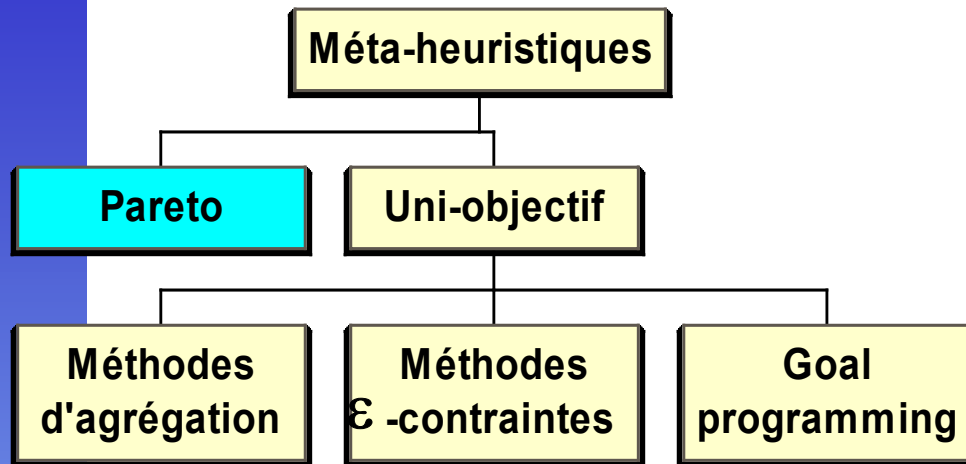
**5 thèses soutenues** : Z. Hafidi, V. Bachelet, D. Kebbal,  
H. Meunier, L. Jourdan

- Problèmes d'optimisation multi-objectifs de grande dimension et complexes dans de nombreux secteurs de l'industrie (**Télécommunications, Génomique, Transport, Environnement, ...**).
- Problèmes de taille sans cesse croissante (explosion combinatoire) **et/ou** Délais de plus en plus courts.

## Optimisation combinatoire multi-objectif

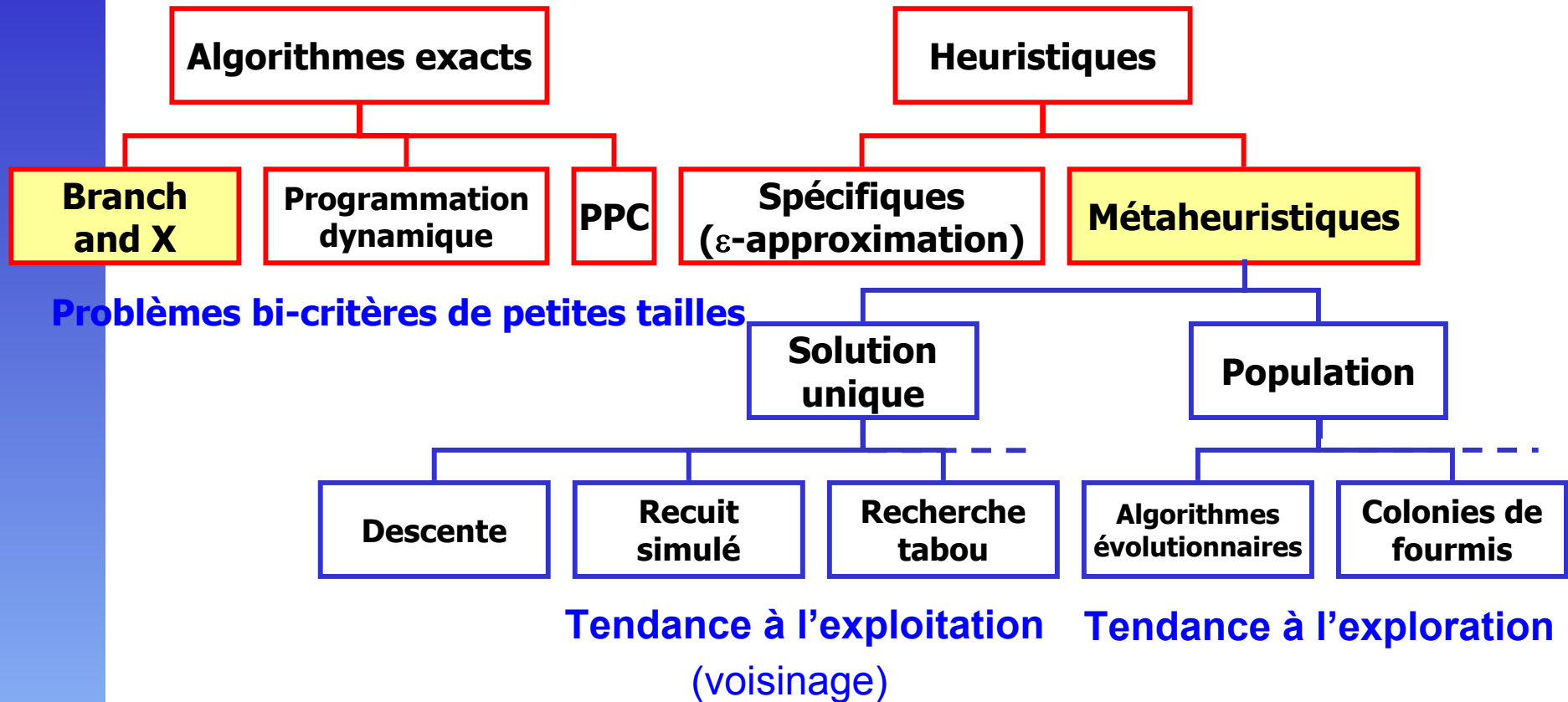
$$\text{(PMO)} \left\{ \begin{array}{l} \min f(x) = (f_1(x), f_2(x), \dots, f_n(x)) \quad n \geq 2 \\ \text{s.c. } x \in S \end{array} \right.$$

- **Objectif du projet DOLPHIN** : Modélisation et Résolution Parallèle de Problèmes d'Optimisation Combinatoire Multi-objectifs de Grandes Tailles



- **Nécessitent des connaissances a priori (poids, contraintes, but)**
- **Modifient la structure du problème**
- **Sensibles au paysage du problème (convexité, discontinuité, ...)**
- **Difficulté : Ordre partiel, Ensemble de solutions, Convexité, ...**

# Méthodes d'optimisation



- **Les méthodes exactes sont vite inutilisables** (problème, instance).
- **Les métaheuristiques sont efficaces** (bornes inférieures, meilleurs résultats).
- **Manque de résultats théoriques** (applicables dans un contexte réaliste).

Résolution de problèmes d'optimisation NP-difficiles

Analyse de paysages

Métaheuristiques hybrides

Conception parallèle de métaheuristiques

Implémentation parallèle et distribuée

Résolution de problèmes d'optimisation multi-objectifs

- Mieux connaître les problèmes et les instances
- Mieux comprendre le comportement des méthodes d'optimisation

Aide à la conception de méthodes parallèles hybrides efficaces

- **Motivation** : Pas de « super » méthode  
**No Free Lunch Theorem** : Wolpert - Macready (1995)
  - ⇒ Introduire dans la méthode une information sur le problème - instance
  - ⇒ Mieux comprendre le comportement des méthodes
- **Structure de l'espace de recherche** : distribution des optima locaux, rugosité, structure de classe de problèmes, ...
- **Résultats** :
  - TSP (opérateurs de recherche),
  - Ordonnancement (fonction objectif),
  - Affectation quadratique (instances),
  - Coloration de graphes (instances, opérateurs, codage).
- **Méthodes auto-adaptatives** : Opérateurs, Méthodes, ...
- **Multi-objectif Pareto** (Agrégation/convexité, Pareto, discontinuité)

• B. Weinberg, E-G. Talbi, « **Fitness Landscape for metaheuristic design** », *Combinatorial Optimization*, 2003.

• E-G. Talbi et al, “**Fitness landscape and performance of metaheuristics**”, *Kluwer Academic Press*, 1999.

# Fronts d'étude



Résolution de problèmes d'optimisation NP-difficiles

Analyse de paysages

Métaheuristiques hybrides

Conception parallèle de métaheuristiques

Implémentation Parallèle et distribuée

Résolution de problèmes d'optimisation multi-objectifs

Hybrider = faire collaborer des comportements opposés (exploration / exploitation)

**Meilleures solutions trouvées pour plusieurs problèmes classiques et réels**

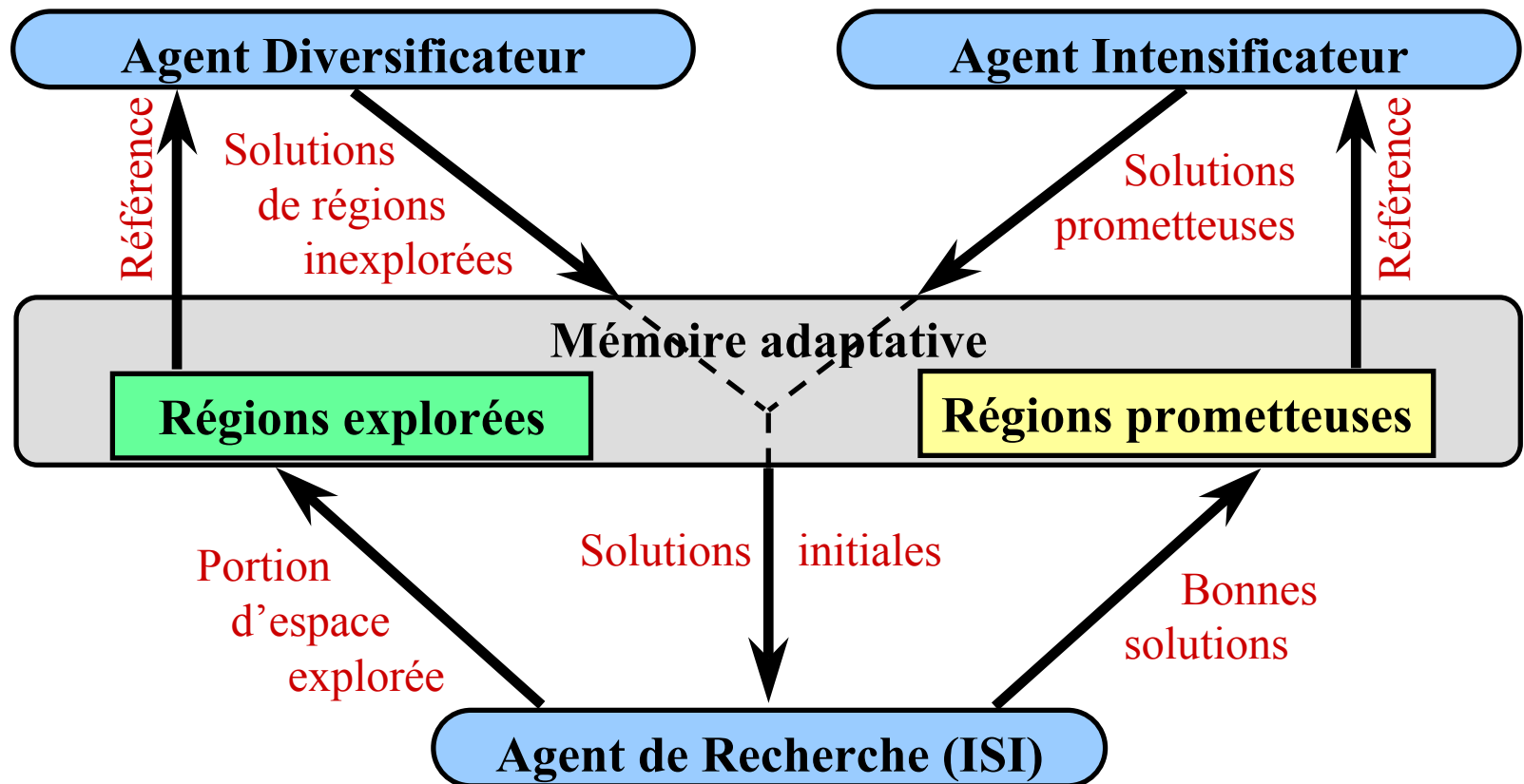
# Méta Hybrides



## ➔ Intérêt de l'hybridation

- Equilibrer exploration / exploitation : Gain de performance, Robustesse
- Taxinomie : Grammaire – Grande variété de méthodes classées

## ➔ CO-SEARCH : Un modèle d'hybridation parallèle





# Coopération Méta-Exacte



Solutions optimales générées par les méthodes exactes

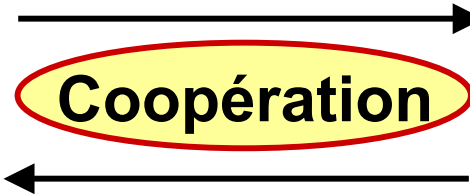
→ Mesurer l'efficacité des méthodes approchées (petites instances)

Bonnes solutions générées par les méthodes approchées

→ Bornes pour l'élagage des branches des méthodes exactes

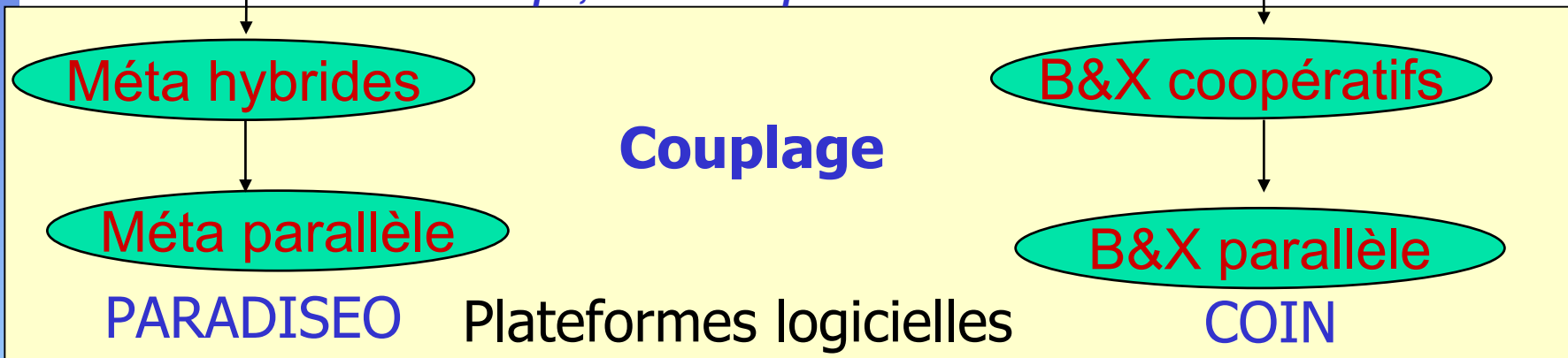
*Réduction de domaines, Upper bound*

Méta  
coopératives  
et parallèles



Branch and  
Bound, Cut  
Multi-critère

*Lower bound, Solution optimale  
sous-pb, Solutions partielles*

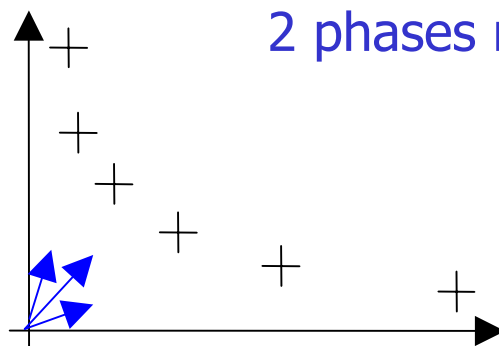


## Exemples de collaboration de solveurs pour l'OC Multicritère

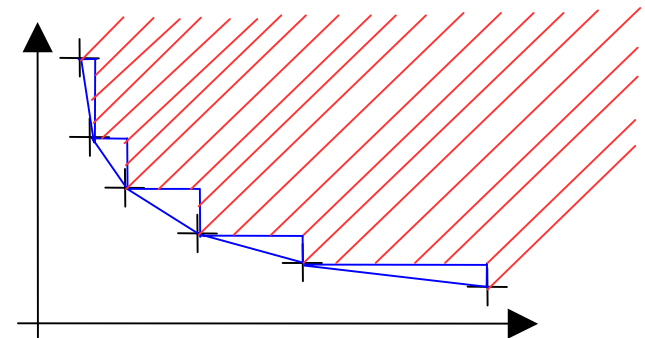
- B&B - AGs pour l'ordonnancement flow-shop
- Branch and Cut & Méta hybride pour le Vehicle Routing Problem :

### Méthode exacte multicritère

- Phase 1 :  
Génération parallèle des solutions supportées  
(un B&B par solution supportée)  
**Aggrégation : Geoffrion's theorem**
- Phase 2 :  
Génération parallèle des solutions non supportées  
(un B&B par triangle)



2 phases recouvrantes



Benchmarks standards : Tai20\*5 : 30 min → Tai20\*10 : 1 semaine

Résolution de problèmes d'optimisation NP-difficiles

Analyse de paysages

Métaheuristiques hybrides

Conception parallèle de métaheuristiques

Implémentation parallèle et distribuée

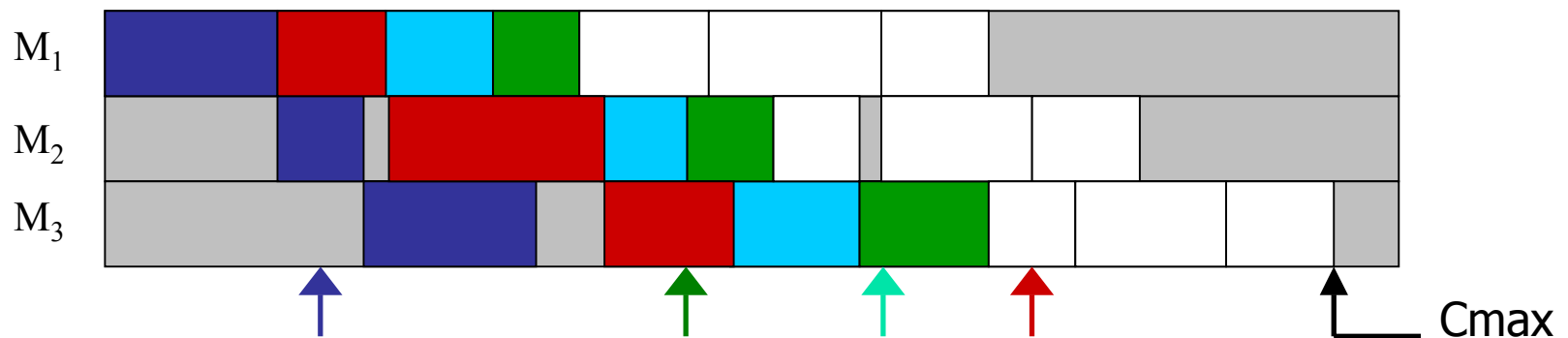
Résolution de problèmes d'optimisation multi-objectifs

- Valider nos approches sur des applications **réelles et génériques**
- Domaines variés : Télécommunications, Génomique.

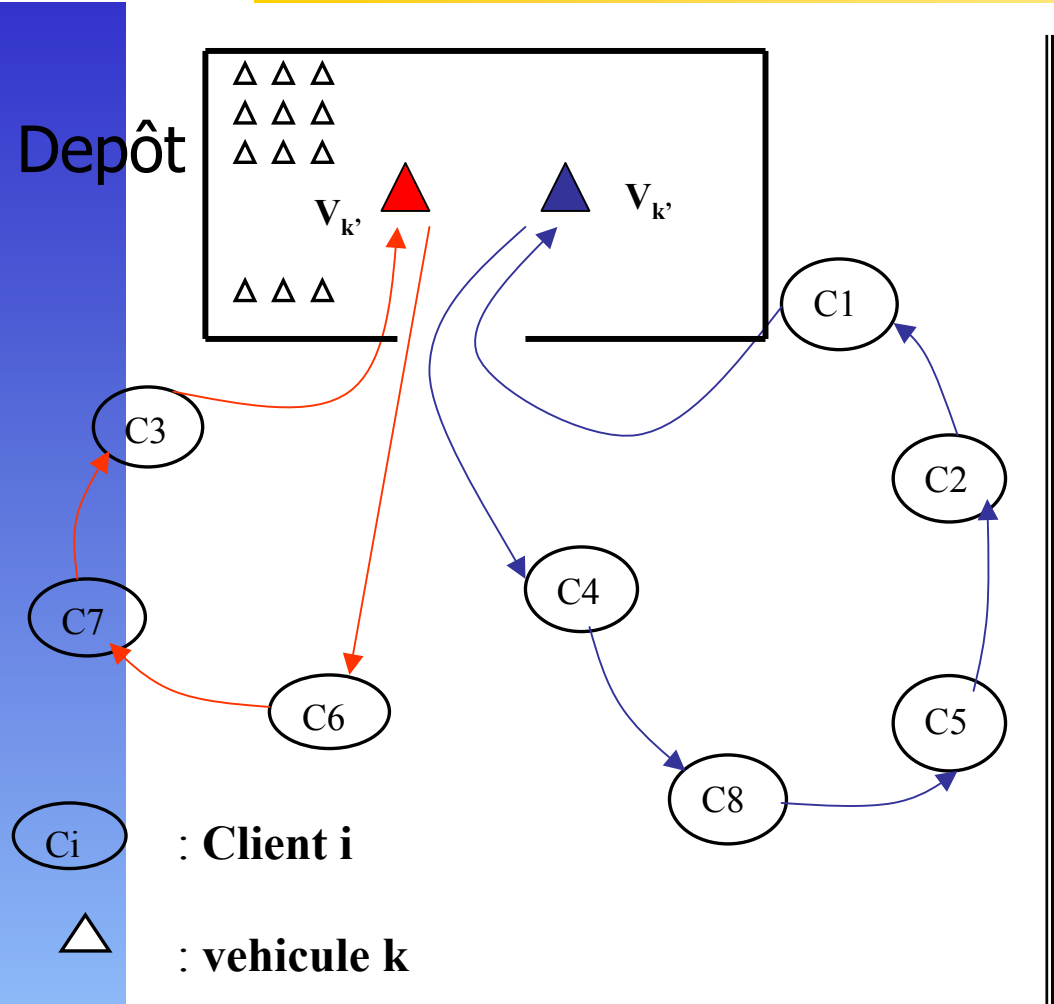
# Ordonnancement flow-shop



- N jobs à ordonnancer sur M machines.
- Flow Shop de permutation.
- Critères optimisés:
  - Cmax : Date de fin d'ordonnancement.
  - T : Somme des retards.
- Problème d'ordonnancement de type F/perm, di/(Cmax,T) [Graham 79].



# VRP : Vehicle Routing Problem



$$I = (G, D, q, Q)$$

$G$  : graphe orienté  $(S, E)$

$D$  : coût des arcs

$q$  : demande des clients

$Q$  : capacité des véhicules

Trouver  $X = (x_{ijk}) \mid x_{ijk} = 1$  si le véhicule  $k$  va du client  $i$  au client  $j$ , 0 sinon

Minimiser 
$$\sum_{i,j=1}^N d_{ij} \sum_{k=1}^M x_{ijk}$$

Contraintes

# VRP : Vehicle Routing Problem



Un problème *naturellement* multi-critère

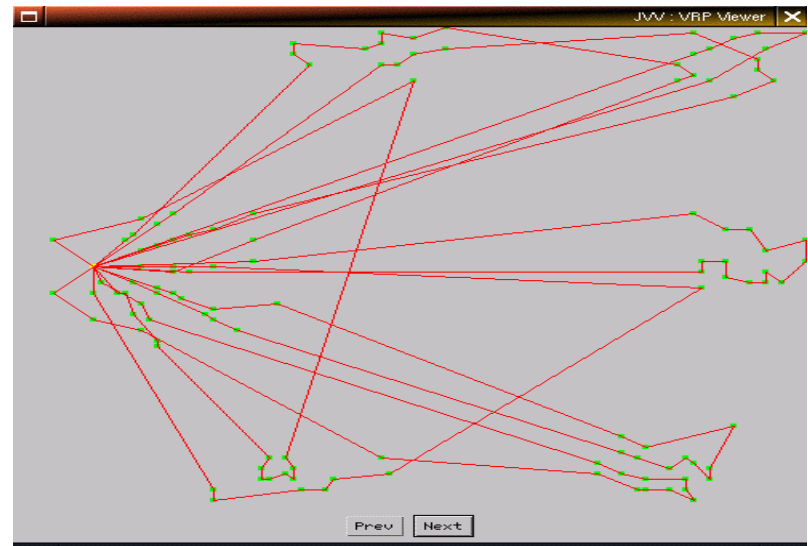
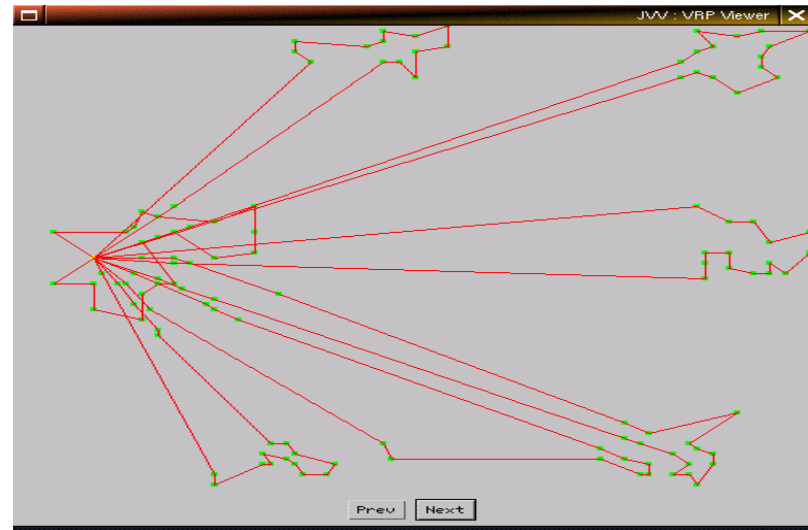
Modélisation bi-critère du problème :

- Premier critère :
  - Minimiser

$$\sum_{i,j=1}^N d_{ij} \sum_{k=1}^M x_{ijk}$$

- Deuxième critère :
  - Equilibrer la longueur des sous-routes
  - Minimiser

$$\max_k \sum_{i,j=1}^N d_{ij} x_{ijk} - \min_{k'} \sum_{i,j=1}^N d_{ij} x_{ijk'}$$



• N. Jozefowicz, E-G. Talbi, F. Semet, "A multi-objective evolutionary algorithm for the covering tour problem", in *Applications of multi-objective evolutionary algorithms*, C.C. Coello, World Scientific, USA, 2004.

# Design d'un réseau cellulaire



(contrat France Telecom, suite projet Esprit ARNO)

**Enjeux économiques** (Coût, Qualité de service)

**Design de réseau** (Positionnement des sites, Paramétrage des antennes)

## ■ Objectifs :

- Min (nombre de sites)
- Min (interférences)
- Min (perte de trafic)
- Max (utilisation des ressources)

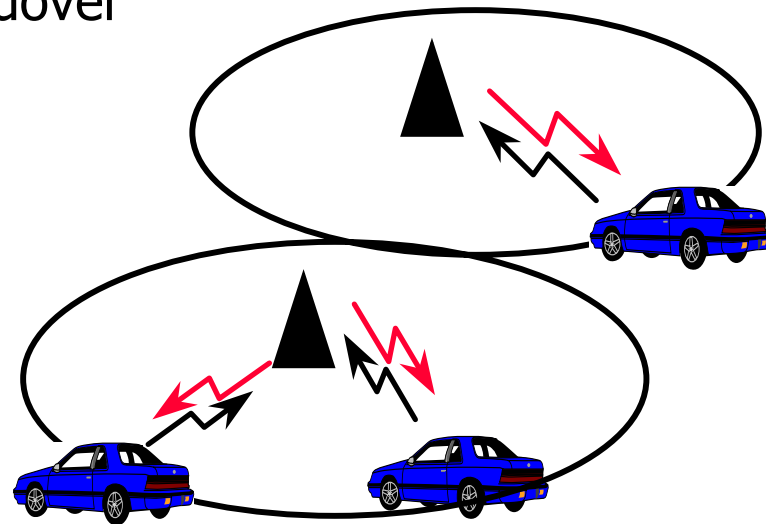
## ■ Contraintes :

- Couverture
- Handover

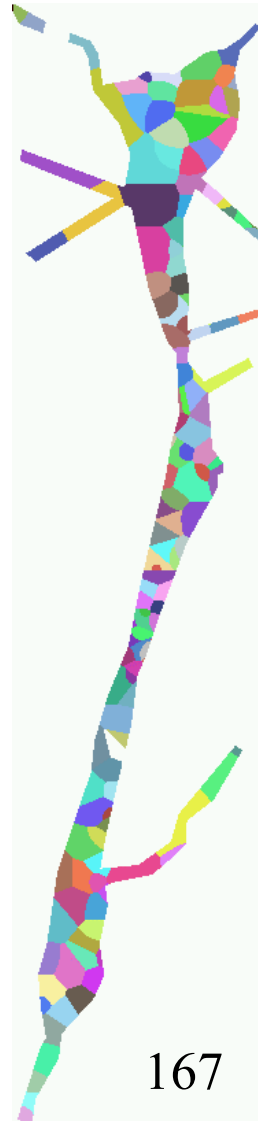
## ■ Espace de recherche :

568 sites candidats →  $2^{3689160}$  solutions et  
~  $600 \cdot 10^9$  choix par antenne !

■ Evaluation coûteuse (pop : 100, Gen :  
 $10^5$ , coût = plus d'un an !!)



# Exemple de réseaux



- Instance : **Autoroute**
- 250 sites, Trafic : 3210 Er





# Applications Génomiques



## Projet GenHom – (2002-2003)

Contexte Post-Génomique (Génopole de Lille - Institut Pasteur Lille, IT-OMICS)

- Règles d'association (tâche de l'extraction de connaissances) pour données génomiques
- Modélisation multicritère du problème
- Application : règles d'associations pour données de puces à ADN :
  - IT-OMICS, GenFit : Maladies cardiovasculaires
  - Institut Pasteur de Lille : Alzheimer
- Hybridation Algorithmes génétiques - méthodes exactes (ex: Apriori)

• L. Jourdan, C. Dhaenens, E-G. Talbi, « **Discovery of genetic and environmental interactions in disease data using evolutionary computation** », *Evolutionary computation in bioinformatics*, Morgan Kauffmann Publishers, USA, 2002.

• M. Khabzaoui, C. Dhaenens, E-G. Talbi, « **Association rules discovery for DNA microarray data** », *Bioinformatics Workshop. of SIAM Data Mining Conf.*, Orlando, USA, 2004.

# Approche multi-critère



**Multiple indicateurs : support, confiance, interet, surprise, conviction, biological criteria, ...**

**Analyse statistique : ACP (Analyse en Composantes Principales)**

[Hilderman et Hamilton, 1999], [Tan et Kumar, 2002], [Adomavicius, 2002], [Lenca et al, 2003]

**Modèle multicritère du problème**

**Design d'algorithmes distribués  
d'optimisation multicritères Pareto**

**Problème d'optimisation  
à grandes échelles**

**Utilisation de la plateforme ParadisEO**

- **Indicateurs de performances :**
  - Contribution, Entropie, Convexité
  - Spacing, ...
- **Elaboration de benchmarks :** MCDM (International Society on Multiple Criteria Decision Making)
  - **Problèmes :** Flow shop, VRP
  - **Résultats :** Meilleurs fronts obtenus, Fronts exacts
- **Logiciel d'analyse et d'évaluation des performances :** GUIMOO (Open Source)

# GUIMOO (A Graphical User Interface for Multi-Objective Optimization)



- Caractérisation de frontières Pareto  
→ Affichage 2/3D  
(mise à jour continue)
  - Structure
    - Continue/Discontinue
    - Concave/Convexe
  - Répartition des solutions dans l'espace des critères

- Analyse des performances  
→ Métriques
  - Contribution
  - Entropie (par niches)
  - Distance générationnelle
  - Spacing

<http://www.lifl.fr/~cahon/guimoo-0.2.tar.gz>

# Evaluations des performances



Guimoo-0.2a

File Visual Metrics Misc.

Radio Network Optimization

Objectives

Label	Aim
Traffic	▲
Number of sites	▼
Interferences	▼

PO\* files

Updating freq. 0

- telecom\_fam1
- telecom\_fam2
- telecom\_fsm1
- telecom\_fsm2

Contents of PO\* files

	Traffic	Number of sites	Interferences
telecom_fam1	2508.53	63	332351
telecom_fam2	1823.54	40	117593
telecom_fsm1	3115.65	128	1.0923e+06
telecom_fsm2	3002.49	74	881638
	3175.05	168	2.61548e+06
	3142.73	122	1.48628e+06

Messages

```
# Loading /home/cahon/Guimoo-0.2/examples/telecom/telecom_fam2 ...
# Processing 203 MO solutions ...
[OK]

# Loading /home/cahon/Guimoo-0.2/examples/telecom/telecom_fsm1 ...
# Processing 224 MO solutions ...
[OK]

# Loading /home/cahon/Guimoo-0.2/examples/telecom/telecom_fsm2 ...
# Processing 167 MO solutions ...
[OK]
```

Gnuplot

Interferences

telecom\_fsm1 (red dots)

telecom\_fsm2 (green dots)

Number of sites (0 to 250)

Traffic (1400 to 3400)

Interferences (0 to 4.5e+06)

Contribution

	telecom_fam1	telecom_fam2	telecom_fsm1	telecom_fsm2
telecom_fam1	0	1	0.85409	0
telecom_fam2	0	0	0.226054	0.527523
telecom_fsm1	0.145907	0.773946	0	0.955357
telecom_fsm2	0.00393701	0.472477	0.0446429	0

Ok

# Evaluations des performances

The screenshot displays the Guimoo-0.2a software interface, which is used for evaluating network performance. The main window shows a 3D visualization of a network structure, likely a tree or a similar topology, with various parameters and constraints.

**PO\* files** window:

- Updating freq. 0
- Files: arno1-1.fit, arno1-2.fit (selected), arno1-3.fit, arno1-4.fit

**Contents of PO\* files** window:

File	Traffic	Number of sites	Interferences
arno1-1.fit	2508.53	63	332351
arno1-2.fit	1823.54	40	117593
arno1-3.fit	3115.65	128	1.0923e+06
arno1-4.fit	3002.49	74	881638
arno1-5.fit	3175.05	168	2.61548e+06
arno1-6.fit	3142.73	122	1.48628e+06
arno1-7.fit	1944.09	48	131292

**Interferences** window:

- Objectives: Traffic, Number of sites, Interferences
- Constraints: Hand Over, Cover

**Cells** window:

- 3D visualization of a network structure with colored cells.

**Gnuplot** window:

- 3D scatter plot showing Interferences vs. Traffic vs. Number of sites.
- Y-axis: Interferences (0 to 4.5e+06)
- X-axis: Traffic (1000 to 3500)
- Z-axis: Number of sites (50 to 150)

**Working Area** window:

- Instance: arno1\_0
- Sites: Number 250, Bounds [3.43e+04, 1.61e+05]
- ZTPs: Number 29954, Bounds [3.64e+04, 1.63e+05]
- Info: Antennas (Omni: 72, Large: 80, Small: 71)

**CFI** window:

- Info: Antennas (Omni: 72, Large: 80, Small: 71)

Résolution de problèmes d'optimisation NP-difficiles

Analyse de paysages

Métaheuristiques hybrides

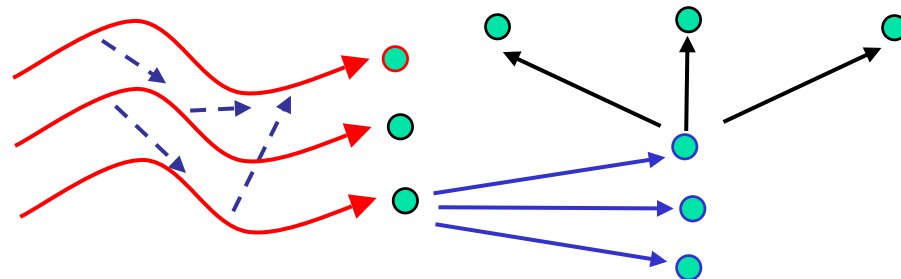
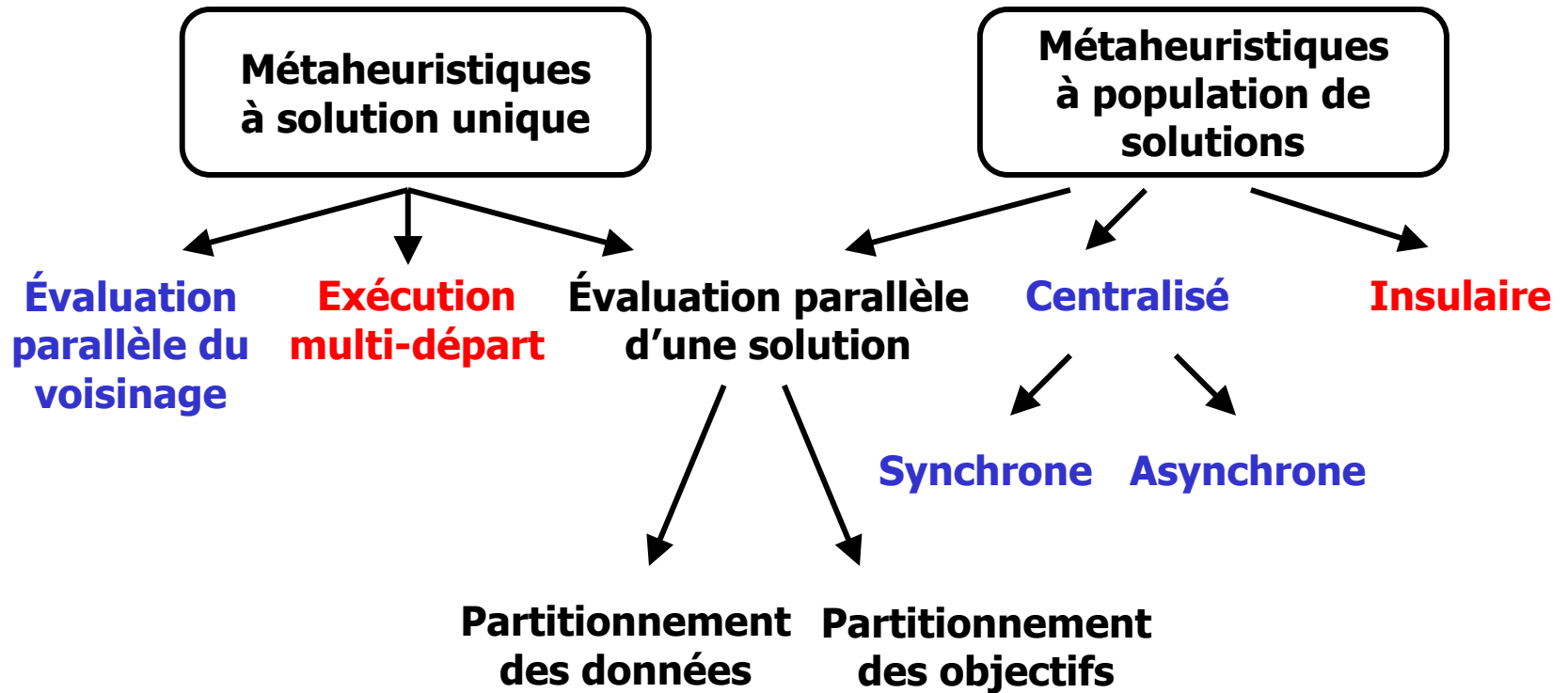
Conception parallèle de métaheuristiques

Implémentation parallèle et distribuée

Résolution de problèmes d'optimisation multi-objectifs

- Accélérer le temps de recherche d'une métaheuristique
- S'attaquer à des problèmes de grandes tailles et complexes
- Améliorer la robustesse

# Modèles parallèles





# PARADISEO (PARAllel and DIStributed Evolving Objects)



EO en quelques mots ... (<http://eodev.sourceforge.net>)

- **Contributions** : Geneura TEAM (Espagne), INRIA (France), LIACS (Pays-bas), LIFL (France)
- **Bibliothèque Open Source C++ «tout objet » (Template)**
- **Indépendante de tout paradigme (métaheuristique)**
- **Flexible / problème traité**
- **Composants génériques (opérateurs de recherche, sélection, remplacement, terminaison, ...)**
- **Nombreux services**
  - **Visualisation, Définition de paramètres en ligne, Sauvegarde/reprise de l'état d'exécution, ...**

---

## → Framework PARADISEO

- **Construire des modèles parallèles et coopératifs dans des contextes hétérogènes (transparence / utilisateur)**
- **Support : Threads, MPI, Athapascan et PVM-Condor**

# Contributions OPAC-LIFL



EO (bibliothèque d'algorithmes évolutionnaires : AG, GP, ES, ...)

Certains mécanismes d'optimisation multicritère

Extension aux algorithmes à base de solutions uniques

ParadisEO (Métaheuristiques parallèles et distribuées)

Méthodes de descente

Recuit simulé

Recherche tabou

Parallélisme

(Partitionnement des solutions, données, recherches, ...)

Coopération

(hybridation synchrone, asynchrone, ...)

Ex. Coopération insulaire

- **Parallélisme** : Accélérer le temps de recherche + S'attaquer à des problèmes de grandes tailles

- **Coopération** : Améliorer la qualité des solutions et la robustesse

# Déploiement sur Clusters & Grilles



## ■ MPICH-G, Posix Threads

### Cluster & SMP Computing

- Communication entre sites géographiquement distribués

## ■ Condor

### Internet Computing

- Tolérance aux fautes, régulation de charge, QoS des propriétaires des stations (Environnement dynamique et imprévisible).
- Reconfiguration dynamique (notification d'événements : panne, préemption, ...).
- Mécanismes de points de reprise mis en place (spécifiques aux métaheuristiques)
- Destiné aux applications indépendantes ou faiblement couplées



## ■ Athapascan (ID-IMAG)

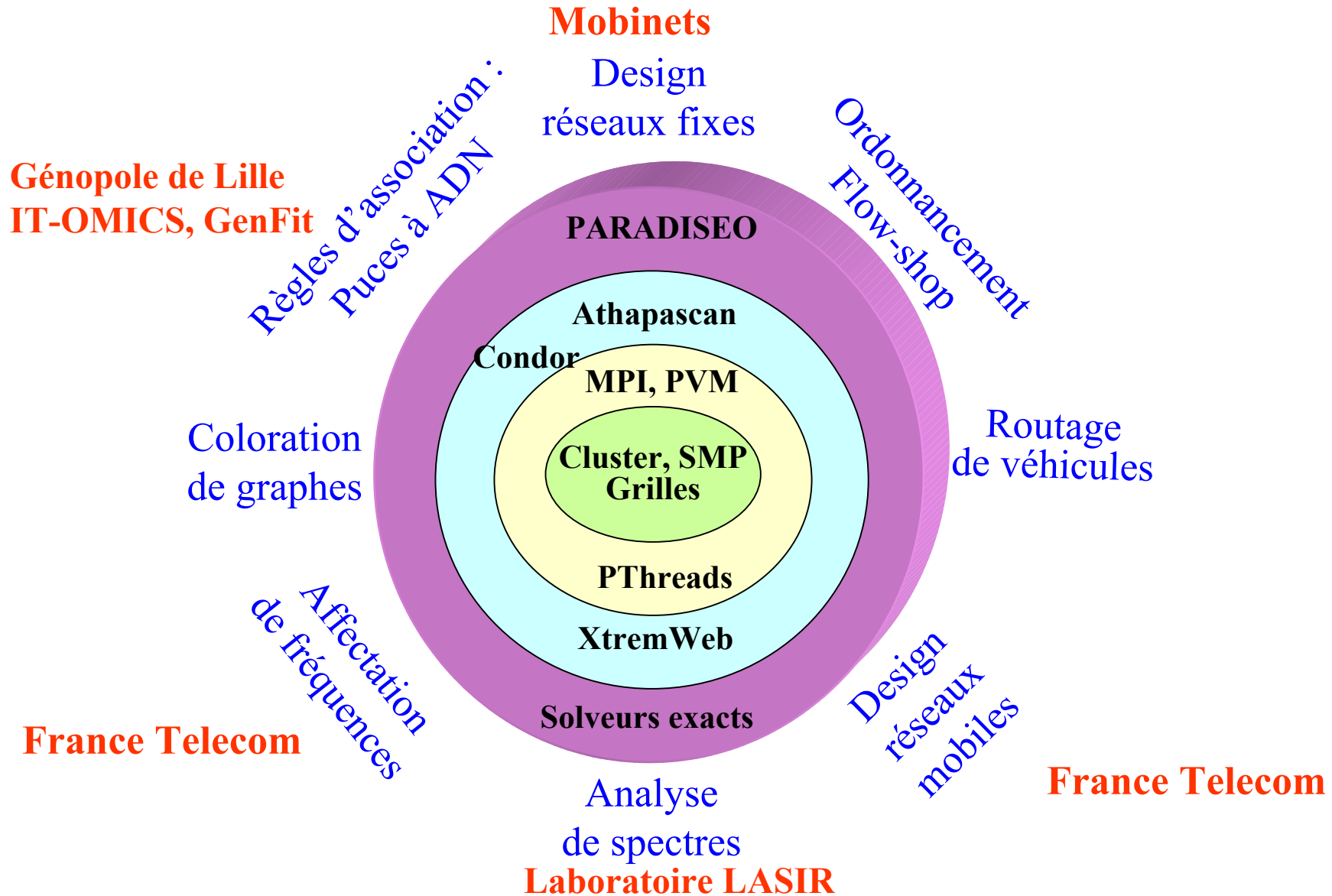
### Virtual Supercomputing

- Ordonnancement (graphe de précédence)

## ■ Peer-to-peer (INRIA Grand Large)

### Internet Computing

# Architecture générale



# Collaborations principales



## ■ Régionales

- **CPER MOST** (Optimisation pour Transports et Télécommunications)
  - ✓ Groupe « Optimisation multi-objectif »
- **Génopole (CIB - Bioinformatique)** – Pasteur, Univ. Lille 2, ...

## ■ CEE

- Réseau EvoNet, EURO-PAREO, EURO-EU/ME
- 6ième PCRD : Grid for Complex Problem Solving - STREPS

## ■ Internationales

- Tunisie, Algérie, Espagne, USA - Illinois

# Collaborations principales



- ACI GRID « DOC-G » (Défis en Optimisation Combinatoire sur Grilles de machines)
  - **Laboratoire ID** (Projet INRIA Apache - Grenoble)
  - **Laboratoire Prism** (Equipe PNN - Versailles)
- ACI GRID «GRID2 » (Groupe de Rencontres, d'Information et de Discussion sur la Globalisation des Ressources Informatiques et des Données:
  - **IRISA, Labri, LRI, LIFL, LIP, ...**
  - **LIFL** : Optimisation discrète, optimisation continue, et extraction de connaissances
  - Ecole d'Hiver Grid'2002 - Aussois  
Journées de travail « Optimisation et extraction de connaissances sur Grilles 2003 » - Lille, ...



## ■ Collaborations projets INRIA

- Projet **Fractales-Complex** (Rocquencourt) : EO – PARADISEO
- Projet **TAO** (Futurs) : ACI IMPBio soumise
- Projet **APACHE** (Rhônes-Alpes) : Athapascan, ACI GRID
- Projet **Grand Large** (Futurs) : XtremWeb

## ■ INRIA complémentaires

- Projet **ALGORILLE** (Lorraine) : Algorithmes pour Grilles
- Projet **MACSI** (Lorraine) : Métaheuristiques - Ordonnancement
- Projet **OPALE** (Rhônes-Alpes et Sophia-Antipolis) : Design multi-critère – Algorithmes évolutionnaires

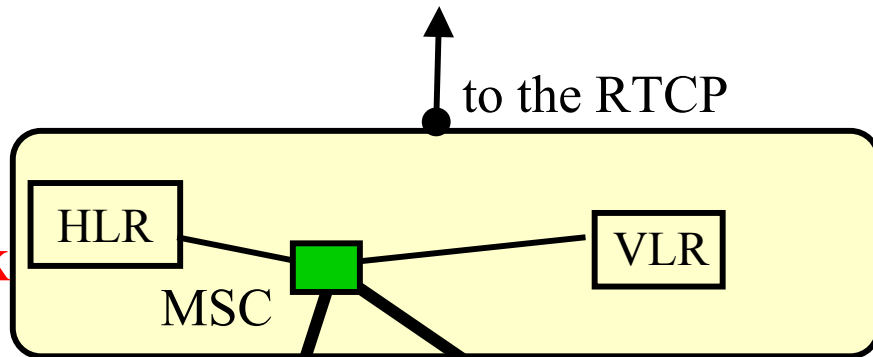
- **Contrat CNET (2000-2002)** - Design de réseaux cellulaires
- **Contrat France Telecom R&D (2004-2006)** - Plateforme logicielle d'optimisation pour les télécommunications
- **Contrat GenFit, IT-OMICS (2002-2003)** : Méthodes d'optimisation pour l'extraction de connaissances en génomique (**Action Bioingénierie – GenHomme**)
- **Mobinets (en cours)** : Design la partie fixe des réseaux de télécommunications mobiles



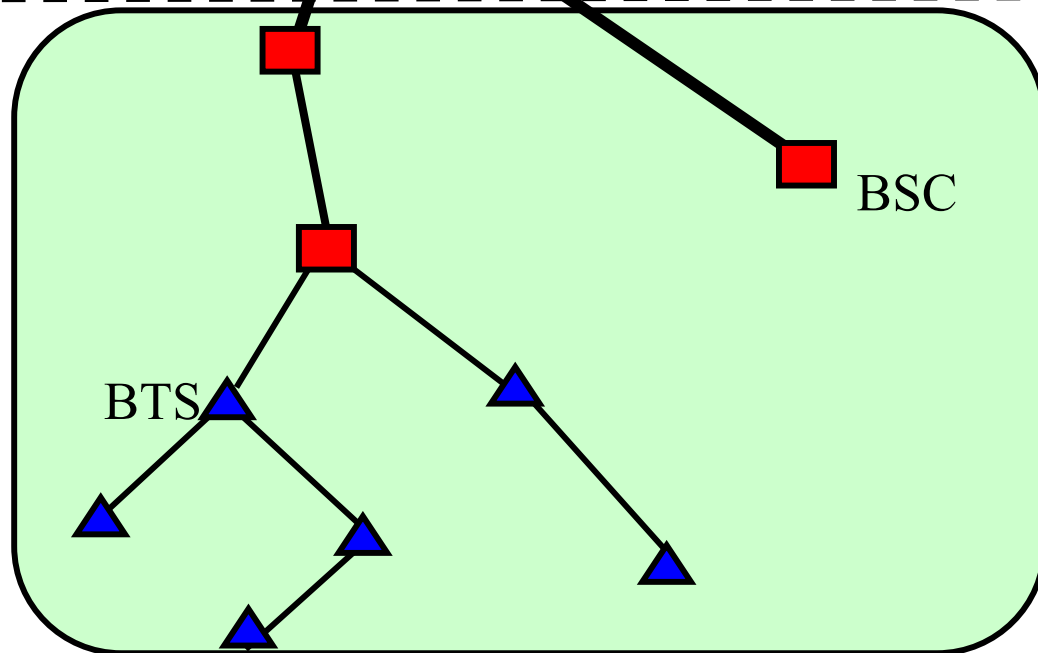
# Architecture réseaux GSM

lifl

Switching sub-network



Radio sub-network



- Objectif : coût, availability
- Contraintes : traffic, capacité, hops, ...

**GSM → UMTS**

- **Opération CNRS Télécom**  
« Modélisation et optimisation multicritère et multipériodique des réseaux de radio communication mobile » (1999-2001)
- **Opération Inter-EPST Bioinformatique**  
« Méthodes d'optimisation pour l'extraction de connaissances en génétique » (2002-2004)
- **AS Action Spécifique du CNRS**  
« Recherche Opérationnelle » (2002-2003)
- **AS Action Spécifique CNRS du RTP 8 GRID**  
« Méthodologie de programmation des Grilles » (2003-2004)
- **AS Action Spécifique CNRS du RTP 8 GRID**  
« Etude préparatoire pour une plateforme de grille expérimentale d'échelle nationale » (2003-2004)

- Comité exécutif de la conférence EMO : Evolutionary Multi-objective optimization
  - EMO'01 : 2001, Zurich – 2003, Portugal – 2005, Mexique
  
- Comité de programme de plusieurs conférences internationales :
  - Métaheuristiques, Algorithmes évolutionnaires : CEC, GECCO, EA, Evocop, JNPC, FRANCORO, ...
  - Parallélisme et distribution : BioSP3 IEEE IPDPS, PPSN, IEEE/ACM CCGrid, ...
  
- Organisation de conférences, workshops et sessions :
  - IEEE IPDPS NIDISC 2002-2004, EuroPar'2000, CEC'99, RenPar'2002, INFORMS'98, ROADEF'2002, MOSIM'2001, PDPTA'97, ROADEF'2003, CESA'2003, Francoro'2004.

- META (Théorie et Application des Métaheuristiques)
  - Animateurs : E-G. Talbi (LIFL), P. Siarry (Paris 12)
  - ROADEF, Inter-GDR : GDR ALP, GDR MACS
  - 130 membres
  
- PM2O (Programmation Mathématique Multi-Objectif)
  - Animateurs : C. Dhaenens (LIFL), V. Gabrel (LAMSADE)
  - ROADEF, GDR I3
  
- JET (Algorithmes Evolutionnaires)
  - Comité de pilotage
  - Organisation Conférence Internationale EA'2005, Lille.

# Autres Utilisateurs et Applications



1. V. Di Martino and M. Mililotti (SORG, Italy) & GGF Scheduling Optimization Scheduling and Resource Management) Chemistry, physics, robotics
2. S. Correia (Chronopost International, France) Vehicle routing
3. M. Schoenauer (INRIA, France) Function optimization
4. J. Eggermont (LIACS, Netherlands) Decision trees, rough sets, NN
5. D. Salle (Lab. Robotique Paris 6), Robotique
6. B. Weinberg (LIFL, France) Frequency Assignment, Coloration de graphes
7. J.J. Merelo (Granada University, Spain) Artificial Life, Neural Networks applied to theoretical physics
8. A. Auger (CERMICS – ENPC, France) Molecular and Multiscale Modelling
9. O. Ricou (EPITA, France) Scientific computation
10. N. Pantelis (TEI Thessaloniki, Greece) Evolutionary algorithms
11. ...

***[http:// eodev.sourceforge.net](http://eodev.sourceforge.net)***