



lip

UPMC
SORBONNE UNIVERSITÉS

Laboratoire d'informatique de Paris 6

UMR 7606

Projet
2014 - 2018



RAPPORT EVALUATION AERES

Projet LIP6 (UMR 7606) - 2014–2018

18 décembre 2012



agence d'évaluation de la recherche
et de l'enseignement supérieur

Table des matières

1	Projet du laboratoire	5
2	Prospectives transverses	15
3	Projet scientifique - Département CALSCI	49
4	Projet scientifique - Département DESIR	63
5	Projet scientifique - Département DAPA	87
6	Projet scientifique - Département NetSys	109
7	Projet scientifique - Département SYSCOMP	123
8	Projet scientifique - Département SOC	141

Chapitre 1

Projet du laboratoire

Sommaire

1.1	Evolution du laboratoire	6
1.2	Projet scientifique	6
1.3	Personnel	10
1.4	Enseignement	11
1.5	Analyse SWOT	12
1.6	Futur organigramme	12

Ce document décrit le projet du laboratoire qui comprend deux volets. Le premier décrit un ensemble de prospectives transverses élaborées par un travail collectif au sein du laboratoire, qui s'est déroulé sur une période de 18 mois en 2011-2012. Cette réflexion s'appuie sur des journées du laboratoire et sur le travail de groupes prospectives thématiques. Le second décrit les perspectives des départements et équipes qui ont été élaborées au sein de chaque département. Le document est composé d'une présentation synthétique du projet, suivie par la présentation des thèmes transverses puis des projets des départements.

1.1 Evolution du laboratoire

L'environnement du laboratoire reste le même, à savoir avec les deux tutelles UPMC et CNRS et deux équipes Inria. Le laboratoire garde la même structure en départements et équipes. Le principal changement concerne la scission du département RSR en deux nouveaux départements "Networks and Systems" (NetSys) et "Systèmes Complexes" (SysComp) qui regroupent respectivement les équipes REGAL, NPA, Phare d'un coté et MoVe, APR et ComplexNetwork de l'autre. Le département NetSys est ciblé sur les aspects réseaux, grands systèmes distribués avec une intersection forte sur les algorithmes répartis entre les équipes NPA et REGAL. M. Potop Butucaru, ex MC de l'équipe REGAL, a été promu en 2012 Professeur et intégrée dans l'équipe NPA pour renforcer cette thématique. Le département SysComp regroupe les thématiques programmation, programmation distribuée, algorithmique des systèmes complexes. Le département RSR était devenu trop important en taille suite à la création des équipes ComplexNetwork et APR, et posait des problèmes d'organisation structurelle et surtout de cohérence des thèmes internes. Cette nouvelle organisation redonne cette cohérence, elle devrait permettre une meilleure animation scientifique et la construction d'un véritable projet au sein de ces nouveaux départements. Par ailleurs, l'équipe MALIRE se scinde en deux équipes LFI (Learning, Fuzzy and Intelligent systems) et MLIA (Machine Learning and Information Access), respectivement centrées sur le traitement de connaissances imparfaites en apprentissage et la conception de systèmes intelligents d'une part et sur l'apprentissage statistique et la recherche d'information d'autre part. Toutes deux restent dans le même département et les collaborations étroites qui ont été développées sur la période précédente continueront. Cette nouvelle structure est destinée à améliorer la visibilité des équipes concernées. Ces évolutions de la structure des équipes et départements ont été décidées en conseil scientifique et en conseil de laboratoire sur proposition des départements et équipes concernées.

1.2 Projet scientifique

1 Elaboration du projet

Le projet du laboratoire sera présenté en deux parties, en premier lieu des perspectives transverses au laboratoire suivies par les perspectives des départements et équipes. Les perspectives transverses constituent des axes novateurs que le laboratoire met en avant. L'exercice pour la laboratoire a consisté à définir de nouvelles directions de recherches, s'appuyant sur le savoir faire du laboratoire et projetant de nouvelles thématiques ou un nouveau déploiement de ces thématiques sur les 5 prochaines années. Les perspectives des départements et équipes sont centrées sur l'évolution des champs spécifiques de recherche propres aux départements et aux équipes.

2 Perspectives transverses

Le travail effectué par le laboratoire pour élaborer sa prospective transverse s'est organisé autour de journées du laboratoire et de groupes de travail thématiques, le tout piloté par une commission d'une dizaine de personnes nommée par le conseil scientifique. Cette commission a été coordonnée par P. Garda (Pr) et par la direction du laboratoire (P. Gallinari, P. Sens). L'ensemble de la réflexion s'est déroulé sur environ 18 mois à compter des premières journées de laboratoire, avec un travail intensif des groupes thématiques sur l'année 2012. À l'issue de deux premières journées de laboratoire, un ensemble de thématiques et de règles pour les construire ont été proposées par la commission au conseil scientifique qui les a amendées et finalisées. L'ensemble a fait l'objet d'une proposition au laboratoire et des groupes de travail ont été constitués. Ces groupes ont été pilotés par un ou deux membres du laboratoire et ouverts à l'ensemble des participants intéressés. Les groupes se sont réunis plusieurs fois. Leurs réflexions ont été exposées et discutées lors de journées de laboratoire et chaque groupe a ensuite produit, sous la responsabilité d'un rédacteur, les documents de prospective transverse qui sont livrés ici. Cette organisation s'est révélée dynamique et efficace. Les groupes de travail ont bien fonctionné, réunissant l'ensemble des forces du laboratoire et suscitant une réflexion et des avancées importantes sur la stratégie du laboratoire et les collaborations entre équipes. Le processus a remporté une adhésion forte au sein du laboratoire.

La construction et le choix des thèmes a pris en compte un ensemble de facteurs, comme les forces et les spécificités en recherche disponibles localement, le contexte régional, national et international, la qualité

novatrice et l'originalité et enfin le réalisme des propositions. Parmi les contraintes définies au départ pour l'exercice, la transversalité a joué un rôle essentiel pour produire des directions de recherche nouvelles qui offrent une alternative et une complémentarité à celles propres aux équipes et qui dépassent les compétences et capacités d'un seul groupe. Le travail réalisé pour l'élaboration des projets de Labex a également été mis à profit lors de cette réflexion et les directions des Labex s'articulent avec celles qui ont été définies pour la prospective. Chaque projet est organisé autour de quelques défis qui constituent les pistes de recherche spécifiques du laboratoire pour le prochain quinquennal. Il s'agit d'un travail véritablement collectif qui représente une vision proposée par l'ensemble du laboratoire.

À l'issue des premières journées du laboratoire, 10 thèmes ont été proposés, et 7 ont été retenus finalement. Ils sont présentés de façon synthétique ci-dessous et détaillés dans la suite du rapport. La description détaillée met en avant le contexte local, les objectifs du laboratoire, le positionnement national/ international, les verrous, la spécificité des axes choisis par le laboratoire ainsi que les pistes et moyens permettant de développer les thèmes. Les 7 projets sont les suivants.

Interactions intelligentes et mobiles. (Rédacteurs M. Diaz de Amorin, A. El Fallah) - Le projet concerne les objets mobiles communicants et l'intelligence ambiante. Il est organisé autour de deux défis. L'un traite des interactions physiques et des infrastructures de communication avec la modélisation de la dynamique des interactions, le développement de composants portables de type radio cognitive et les solutions distribuées pour le transfert et le stockage de données. L'autre, plus axé sur les aspects algorithmiques, porte sur l'intelligence ambiante dans des systèmes à grande échelle, avec la modélisation et la simulation multi-échelles, l'algorithmique répartie, l'analyse comportementale des interactions et la vérification, la distribution des calculs sur les capteurs intelligents et les différents éléments des systèmes. Les thématiques abordées concernent les réseaux, les systèmes et les logiciels distribués ainsi que l'intelligence artificielle.

Fiabilité/ Sécurité. (Rédacteur J.C. Faugère) - Le projet traite des enjeux de fiabilité dans les systèmes informatiques (QoS, sûreté de fonctionnement, tolérance aux fautes) et de sécurité (protection, détection de fuites). Un premier défi traite de l'analyse et de la maîtrise/protection des fuites physiques dans les composants cryptographiques, de la mise en œuvre d'attaques combinées physiques/algébriques et d'approches formelles pour la mesure de fuites et la détection d'attaques potentielles. Les approches proposées combinent les aspects matériels et théoriques (modélisation par automates). Le second défi concerne la fiabilité des logiciels pour les systèmes répartis à travers la définition de langages de haut niveau permettant d'améliorer cette fiabilité sur la chaîne complète : conception, programmation, exécution, maintenance. Ce défi concerne les thématiques algorithmique et logiciel, calcul formel, systèmes embarqués.

Traitement distribué de données distribuées. (Rédacteur T. Artières) - Il s'agit d'un projet "Big Data" ciblé à la fois sur les algorithmes pour l'analyse des données massives et les traitements répartis. Un premier défi concerne le passage à l'échelle des algorithmes dans un cadre de traitements centralisés avec comme objectif la conception d'algorithmes sous linéaires ou d'algorithmes à budget limité. Un second défi concerne les traitements distribués et les données dynamiques et distribuées. Actuellement, les deux aspects distribution des traitements et des données sont rarement traités simultanément. Le projet fait intervenir les thématiques bases de données, apprentissage, réseaux et systèmes.

Nouvelles approches en optimisation (Rédacteur P. Perny) - L'optimisation est présente dans une grande partie des problématiques de recherche du laboratoire, le projet vise à développer et diffuser au sein du laboratoire des techniques modernes d'optimisation et à faire émerger de nouvelles approches par l'analyse de problématiques pratiques. Deux défis sont proposés. Le premier traite d'optimisation sous contrainte d'espace pour les problèmes de grande taille et d'optimisation dynamique par des algorithmes en ligne par exemple. Le second défi concerne l'optimisation pour des objectifs multiples reflétant divers points de vue et l'optimisation avec garantie de performances. Les thématiques abordées sont l'algorithmique, les réseaux, la décision, la recherche opérationnelle.

Nouveaux Systèmes Répartis. (Rédacteur P. Sens) - Le projet traite des architectures réparties dans les grands systèmes (e.g., les nuages informatiques) et dans les architectures multi-cœurs. Un premier défi concerne

l'étude des grands systèmes dynamiques avec l'analyse qualitative et quantitative de ces environnements et la définition de protocoles et méthodes de conception pour développer des services. Le second défi traite des architectures multi-cœurs avec un accent mis sur les problématiques de passage à l'échelle et d'accès à une mémoire partagée non-uniforme. Ce projet concerne les thématiques programmation, systèmes répartis, calcul scientifique et systèmes embarqués.

Santé. (Rédacteur P. Garda) - Parmi les enjeux de société évoqués dans la construction de la prospective, nous avons retenu une thématique liée à la santé, qui bénéficie de l'implication de plusieurs équipes du laboratoire et du contexte local à l'UPMC. Deux défis sont abordés ici. Le premier concerne l'analyse et l'interprétation des données de la santé avec la modélisation des émotions et des comportements ainsi que la conception de capteurs intelligents. Le second concerne la e-santé individualisée et se caractérise comme un domaine d'application des méthodes et technologies développées dans le premier défi. Trois axes avec un fort enjeu sociétal sont envisagés : l'assistance à la personne, l'autonomie et le maintien à domicile, la santé personnalisée. Les thématiques concernées sont l'intelligence artificielle, l'algorithmique, les systèmes embarqués, le traitement d'images.

Social Computing. (Rédacteur L. Denoyer) - L'analyse de grands réseaux sociaux soulève des problèmes trans-disciplinaires et transversaux au laboratoire. Le projet se focalise sur l'étude de la dynamique de ces réseaux sous l'angle de deux défis : la dynamique structurelle (évolution des communautés, des structures, etc.) et la dynamique des contenus diffusés sur ces réseaux qui comprend aussi bien la conception de modèles explicatifs que de modèles permettant d'agir sur la diffusion. Il attaque trois problèmes clés de cette problématique de la dynamique qui sont la mesure, la modélisation et l'analyse de l'interaction. Le projet est pluri-disciplinaire et repose sur des collaborations extérieures au laboratoire. Il présente l'opportunité de fédérer une recherche pluri-disciplinaire en Ile de France. Il concerne au LIP6 les thématiques apprentissage, bases de données, multi-agents, algorithmique et systèmes complexes.

Liens avec le labex SMART. Sur ces 7 thèmes prospectifs, 5 recourent largement les axes du Labex SMART sur lequel le laboratoire est engagé. Nous listons ci-dessous les 5 axes de SMART en indiquant les relations de chacun avec les thèmes transverses du laboratoire.

1. La compréhension de l'humain dans ses dimensions cognitives, neurophysiologiques comme biomécaniques.
2. Les interfaces et l'interaction de l'humain avec des environnements numériques et des mondes physiques distants.
3. Le développement des services numériques pour l'accès à la connaissance et à l'information, le traitement des données numériques, qui recoupe les thèmes transverses *Interactions intelligentes et mobiles*, *Traitement distribué de données distribuées* et *Social Computing*.
4. La mobilité dans les réseaux et l'interopérabilité, qui recoupe les thèmes transverses *Interactions intelligentes et mobiles* et *Nouveaux Systèmes Répartis*.
5. La création de services et de technologies associées aux besoins de l'e-santé qui recoupe le thème transverse *Santé*.

Par ailleurs, les thèmes transverses *Interactions intelligentes et mobiles*, *Traitement distribué de données distribuées* et *Social Computing*, *Santé* sont en lien avec le plan d'orientation du pôle de compétitivité Cap Digital dans lequel le laboratoire est largement impliqué.

3 La prospective des départements

Equipes et départements ont également élaboré des perspectives axées sur leurs thématiques de recherche propres qui sont décrites dans la deuxième partie de ce document prospective. Nous en donnons ci-dessous une brève synthèse.

CALSCI. Le département CALSCI sera constitué des équipes PEQUAN et PolSys (équipe commune UPMC, CNRS, Inria) comme dans la configuration actuelle. Il développe deux thématiques transverses aux équipes, le calcul symbolique/ numérique et la cryptographie embarquée. L'équipe PEQUAN devrait voir son activité HPC augmenter avec la participation à l'Institut du Calcul et de la Simulation de l'UPMC et au Labex CalSimLab qui porte sur la modélisation et la simulation numérique en chimie et biologie. Une thématique sur les opérateurs pour la cryptographie renforce les liens avec PolSys. Cette dernière équipe a recentré ses activités sur le calcul et sur l'utilisation des bases de Gröbner. Le département est impliqué dans les thématiques transverses fiabilité/sécurité, nouvelles approches en optimisation, nouveaux systèmes répartis, santé.

DAPA. Le département sera constitué des équipes LFI (Learning, Fuzzy and Intelligent systems), MLIA (Machine Learning and Information Access), BD et ACASA. Les deux premières équipes sont issues de l'actuelle équipe MALIRE. Les recherches sont centrées sur l'apprentissage automatique et la gestion de grandes masses de données hétérogènes, réparties et dynamiques (flux). Le projet pour les 5 prochaines années vise des avancées théoriques sur la conception de nouvelles techniques et d'algorithmes d'apprentissage et de traitement de données pour l'accès à l'information dans un environnement complexe. 5 défis sont mis en avant, *Ultra Large Scale* qui concerne le passage à l'échelle, *Données distribuées* pour les aspects bases de données, *Traitement atomique de données distribuées* concerne la conception d'algorithmes d'apprentissage dans un cadre réparti, *Dynamisme* concerne le traitement de données en flux et évolutives, *Sémantique/ extraction* s'attache au traitement des données sémantiques textuelles et multi-média. Le département s'implique dans les axes transverses traitement distribué de données distribuées, santé et social computing.

DESIR. Le département sera constitué comme actuellement des équipes RO, DECISION, SMA et MOCAH. Les recherches portent sur la modélisation et l'optimisation des systèmes, l'aide à la décision et la décision automatique, les agents adaptatifs et les systèmes multi-agents, les systèmes interactifs et de formation. 7 défis sont mis en avant qui concernent les domaines suivants : *Nouveaux défis pour l'ordonnancement*, *Programmation mathématique* pour les algorithmes en ligne et l'approximation, *Optimisation combinatoire multicritère* pour l'identification et l'approximation de l'ensemble de Pareto, *Décision collective, coordination et mechanism design* pour les systèmes multi-agents, *Modèles graphiques* pour le raisonnement et la décision, *Jeux vidéo et jeux sérieux* avec la définition de stratégies d'apprentissage basées sur la coopération et la compétition dans les jeux, *Modélisation de l'utilisateur* pour adapter les systèmes informatiques à l'humain. Le département est impliqué dans les axes transverses interactions intelligentes et mobiles, nouvelles approches en optimisation, santé et social computing.

NetSys. Le département regroupe les équipes NPA, Phare et REGAL de l'actuel département RSR. Les activités de recherche concernent les réseaux (NPA, Phare) et les systèmes distribués (NPA, REGAL), avec plus particulièrement les thèmes suivants : calcul distribué, systèmes P2P, réseaux mobiles et ambiants, grands systèmes et cloud, tolérance aux fautes. 3 défis sont mis en avant, la *concordance des modèles vis-à-vis des environnements réels*, *l'adéquation des algorithmes vis-à-vis des problèmes à résoudre et des tâches correspondantes*, la *conception de solutions auto-organisées et capables de s'adapter aux changements d'échelle*. Le département est impliqué dans les axes transverses fiabilité/sécurité, traitement distribué de données distribuées, nouveaux systèmes répartis.

SysComp. Le département regroupe les équipes APR, MoVe et ComplexNetworks de l'actuel département RSR. La thématique générale est celle de l'étude des grands systèmes avec des entités en relation les unes avec les autres. Elle est abordée sous les aspects algorithmique (APR), modélisation (ComplexNetworks), conception et vérification (MoVe). Les défis avancés sont les suivants : *Modélisation basée sur les mesures* avec le déploiement de procédés de mesure et la validation de modèles, *Vers plus de formalisation*, *Gérer les grandes tailles* au niveau algorithmique avec la maîtrise de la complexité en espace et les représentations compactes, *Dynamiques et hétérogénéités* pour la description et l'analyse des grands systèmes dynamiques. Le département s'investit dans les thèmes transverses interactions intelligentes et mobiles, fiabilité/sécurité, nouvelles approches en optimisation, nouveaux systèmes répartis, social computing.

SOC. Le département conserve sa structure actuelle avec les équipes ALSOC, CIAN et SYEL. La thématique de recherche générale concerne la conception de circuits intégrés sur puce avec une vocation applicative. Les défis proposés sont les suivants : *Systèmes hétérogènes* avec le développement d'outils pour la conception analogique-numérique, *Fiabilité et sécurité des circuits et systèmes* pour l'évaluation de leur robustesse vis à vis d'un ensemble de critères, *Architectures manycore* avec le développement d'outils de conception de ces architectures et leur optimisation, *Architectures reconfigurables* dynamiquement et optimisées pour la consommation et l'utilisation de ressources, *Technologies émergentes* (bionique, photonique, etc.) pour la conception de nouvelles architectures, *Systèmes embarqués pour la santé* en lien avec le secteur médical. Ces activités sont liées aux thèmes transverses interactions intelligentes et mobiles, fiabilité/ sécurité, nouveaux systèmes répartis et santé.

1.3 Personnel

1 Académique

Le laboratoire a bénéficié de nombreux départs à la retraite pendant les 8 années qui viennent de s'écouler. De nombreux départs concernaient des E/C UPMC en 27ème section hors laboratoire et le LIP6 a pu ainsi s'enrichir de nombreux E/C. Cette période est terminée et ce sont actuellement de jeunes E/C qui quittent le laboratoire sur des promotions. Ces départs concernent les forces vives du laboratoire, essentielles au fonctionnement des équipes. Ces jeunes chercheurs confirmés sont souvent porteurs eux mêmes de thématiques de recherche originales essentielles pour leur équipe. Cet essaimage constitue une opportunité à long terme car elle porte les germes de nouvelles collaborations avec d'autres laboratoires, elle constitue également un défi pour renouveler ces forces vives, en particulier dans cette période de restriction de ressources humaines. Ce défi comporte deux points sensibles. Le premier est le renouvellement de ces effectifs stratégiques qui devra être effectué en discussion avec l'UPMC. Le second concerne la politique de recrutement qui devra offrir un bon compromis entre recrutements externes et promotions internes basées sur l'excellence des candidats.

Concernant les chercheurs des organismes de recherche, le laboratoire a fait sur cette dernière période d'importants progrès en recrutant 5 CR CNRS, 1 DR CNRS, 2 DR Inria avec 1 CR CNRS et 1 CR Inria supplémentaires en octobre 2012. Il a par ailleurs obtenu 2 promotions CR -> DR CNRS. Sur la période précédente, seul 1 chercheur avait été recruté. Cette progression est due à la politique incitative du laboratoire, visant à la sensibilisation et à la mobilisation des équipes pour attirer des candidats de valeur sur des postes de CR, qui passe souvent par l'accueil préalable de post-docs, et à une politique pro-active de l'INS2I. Même si ces recrutements ont permis de renforcer plusieurs équipes, le nombre de chercheurs reste globalement faible avec 22 chercheurs pour 150 E/C. Il reste encore une marge de progression importante pour attirer de jeunes talents qui pourrait passer par une véritable politique globale au laboratoire sur le recrutement de post-docs et par une meilleure communication des opportunités de recrutement. C'est le deuxième défi qui attend le laboratoire sur cette nouvelle période.

L'accueil de professeurs et chercheurs invités passe principalement par les mois d'accueil mis à disposition par l'UPMC, le CNRS, et des programme bilatéraux, principalement pour de courtes périodes. Inria offre aux équipes concernées des accueils sur des durées plus longues mais qui restent limitées à 2 équipes pour le LIP6. Quelques opportunités permettant des échanges plus ambitieux comme les ANR internationales (cf. accueil de S. Russell Pr Berkeley en 2012 - 2013). L'accueil de chercheurs seniors sur des périodes longues est un enjeu important pour lequel le laboratoire doit trouver des solutions. La participation aux Labex, en particulier SMART, offre une première piste pour cela.

2 Administratif et technique

L'organisation administrative et technique a largement évolué en quelques années avec aujourd'hui, une responsable - secrétaire générale - du laboratoire (C. Mercier) et un responsable technique (C. Boudier) pour coordonner le travail des ingénieurs. Pour l'administration, le modèle que nous mettons progressivement en place consiste à confier la gestion complète administrative et financière d'un ensemble d'équipes à un administratif, ce qui inclut en particulier la gestion des personnels permanents et temporaires et des contrats et conventions. Cette organisation permet aux équipes d'avoir un contact unique pour l'ensemble des problèmes administratifs

et offre une modularité raisonnable à l'occasion des mouvements de postes et remplacements. Pour le personnel, elle permet une activité plus enrichissante qu'une fragmentation des tâches et potentiellement de meilleures opportunités de promotion. Coté technique, le service informatique était organisé sur la période précédente autour de pôles d'ingénieurs chargés de la gestion des départements et de fonctions communes. Il est actuellement en réorganisation sous l'égide de son responsable afin de mieux mutualiser les ressources et services au niveau du laboratoire. Les services de proximité sont conservés sur la base d'une répartition des ingénieurs au sein des équipes, un travail important est effectué sur l'uniformisation des services communs, des pratiques et des services aux équipes. Des groupes de travail sont constitués autour de missions de service et de l'administration d'équipements transverses au laboratoire. Au delà d'une uniformisation nécessaire des pratiques, cette nouvelle organisation doit offrir un caractère plus attractif pour les missions des ingénieurs et aider à l'évolution de leur carrière. Un des objectifs de ce travail est également de permettre par une meilleure organisation de libérer du temps ingénieur pour des mises à disposition sur des charges de support recherche aux équipes. Cet enjeu important a jusqu'à présent été freiné par la charge de travail liée à l'accroissement des besoins utilisateurs, à la mise au point de nouvelles infrastructures et aux déménagements du laboratoire (4 sur 10 ans) qui nécessitent à chaque fois une réorganisation des infrastructures.

En ce qui concerne les effectifs, le laboratoire a un nombre relativement faible de supports administratifs et techniques et chaque mouvement de personnel nous fait basculer dans une situation délicate à gérer, qui passe par le recrutement et le financement de collaborateurs administratifs et techniques sur des postes temporaires. En septembre/octobre 2012, nous faisons face à 3 départs (1 IE systèmes-réseaux, 1 T gestionnaire financier, 1 T électronique) et 2 départs de gestionnaires financiers sont prévus fin 2013 qui nous mettront dans une situation extrêmement délicate pour employer un euphémisme. La communication est un autre point faible du laboratoire qui devra être améliorée. C'est un élément essentiel pour la gestion et la reconnaissance d'un laboratoire de cette taille. Elle passe par le recrutement d'agents dédiés à cette tâche. La responsable de la communication qui occupait un poste d'IE CNRS étant partie en retraite en juin 2012, le service est actuellement réduit à une seule personne spécialisée sur le graphisme. Le laboratoire ne pourra pas construire un service de communication sans remplacement de ce poste. Le renouvellement de l'ensemble de ces postes par les tutelles est un enjeu crucial pour l'avenir du laboratoire.

1.4 Enseignement

Le projet d'enseignement prévoit de conserver une mention informatique avec un ensemble de spécialités qui évolue sensiblement. La nouvelle organisation du master est en meilleure adéquation que l'actuelle avec la structure du laboratoire et offre un accès mieux équilibré à toutes les équipes aux enseignements de master. La nouvelle maquette prévoit 8 spécialités. Elles partagent de nombreux cours notamment en M1 et la notion de spécialité liée à la recherche prend véritablement son sens en M2. Nous donnons ci-dessous les correspondances spécialité de master - département du LIP6.

Liste des spécialités du master informatique et correspondance avec les départements du LIP6

- ANDROIDE - AgeNts Distribués, Robotique, Recherche Opérationnelle, Interaction, DEcision correspond au département DESIR.
- DAC - Données Apprentissage et Connaissances, correspond au département DAPA
- IMA - IImage, regroupe des compétences CALSCI et DAPA
- RES - RESeaux est assuré par des équipes des département NetSys et SysComp
- SAR - Systèmes et Applications Réparties, est supportée par les départements SysComp et NetSys
- SESI - Systèmes Electroniques, Systèmes Informatiques correspond au département SOC
- SFPM - Sécurité, Fiabilité et Performance du Numérique correspond au département CALSCI
- STL - Sciences et Technologies du Logiciel, est assurée principalement par des équipes du département SysComp

Un enjeu important pour l'informatique est le développement des enseignements internationaux. Plusieurs spécialités dans la nouvelle maquette proposent des cours en anglais. Les initiatives comme les masters internationaux DMKM (Erasmus Mundus) et réseaux (Vietnam) devront faire l'objet de demandes de renouvellement ou de prolongation. Ces initiatives doivent être intensifiées en lien avec les relations internationales de l'UPMC.

Par ailleurs, l'Ecole doctorale Edite, à laquelle les équipes du LIP6 sont affiliées, est reconduite dans sa configuration actuelle, même si elle doit évoluer à terme avec le départ programmé à Saclay du partenaire Télécom ParisTech qui est également un des partenaires principaux du LIP6. Les actions en lien avec l'EIT ICT Labs vont voir un renforcement du programme "Doctoral School on ICT Innovation" qui propose aux étudiants de l'EDITE des périodes de formation à l'innovation (actuellement 3 mois avant et 3 mois après la thèse). En particulier le nombre de jetons disponibles pour l'EDITE va être augmenté.

1.5 Analyse SWOT

Points forts :

- large couverture de thématiques informatiques théoriques et appliquées
- visibilité externe du laboratoire et de l'UPMC
- production scientifique de bonne qualité avec certaines équipes à un niveau excellent

Points faibles :

- taille du laboratoire par rapport au support recherche disponible (administratif et ingénieur), avec en particulier l'absence d'ingénieurs support recherche permanents
- communication globalement à améliorer - absence de ressource humaine pour ce poste
- départs massifs d'administratifs (retraite etc.)
- possibilité d'attirer de jeunes talents ou des seniors renommés et de leur offrir des postes
- relations internationales encore peu structurées au niveau laboratoire et enseignement de master
- manque de ressources pour des invitations longue durée régulières de chercheurs étrangers
- charges liées au fonctionnement de la structure recherche/enseignement qui mobilisent une partie trop importante des E/C et chercheurs du laboratoire

Risques liés au contexte :

- capacité matérielle à réunir financements et ressources pour les objectifs fixés en particulier pour les thématiques transverses qui ont été élaborées
- départs importants de jeunes E/C confirmés sur promotion et incertitude sur leur remplacement
- attractivité du laboratoire dans un contexte international - perte d'attractivité des postes E/ C
- recrutement d'étudiants de qualité dans les masters et en doctorat

Possibilités liées au contexte :

- développement de collaborations pluridisciplinaires dans le cadre UPMC et Sorbonne Universités
- implication dans un ensemble conséquent de projets investissement d'avenir
- relations avec de nombreux acteurs industriels et avec le pôle de compétitivité Cap Digital

1.6 Futur organigramme

Les figures 1.1 et 1.2 donnent le futur organigramme du laboratoire.



FIGURE 1.1 – Organigramme du laboratoire : future organisation des départements

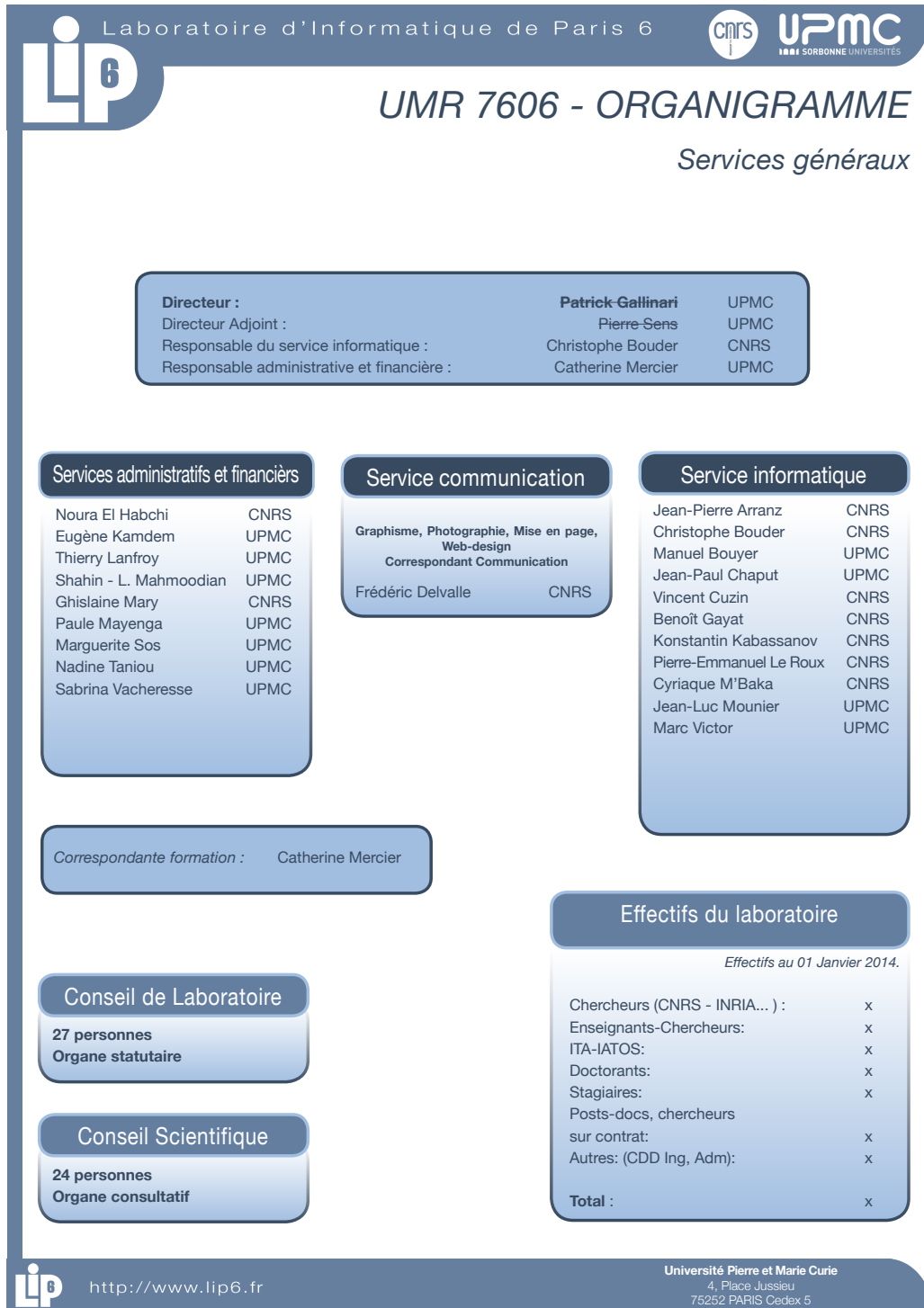


FIGURE 1.2 – Organigramme du laboratoire : future organisation des services

Chapitre 2

Prospectives transverses

Sommaire

2.1	Projet Transverse 2014 – 2018 - Thème AmI-oc	16
2.2	Projet Transverse 2014 – 2018 - Thème Fiabilité/Sécurité	20
2.3	Projet Transverse 2014 – 2018 - Thème LSDP	27
2.4	Projet Transverse 2014 – 2018 - Thème Optimisation	31
2.5	Projet Transverse 2014 – 2018 - Thème NSR	35
2.6	Projet Transverse 2014 – 2018 - Thème Santé	40
2.7	Projet Transverse 2014 – 2018 - Thème SocComp	44



Vague D : Campagne d'évaluation 2012-2013

2.1 Projet Transverse 2014 – 2018 - Thème AmI-oc

1 Présentation du thème “Interactions Intelligentes et Mobiles”

Le changement du comportement des usagers face à l'évolution des systèmes informatiques communicants, ainsi que l'apparition d'équipements plus performants et de dimensions réduites, modifient peu à peu la façon dont les environnements dans lesquels les usagers sont immergés doivent opérer. Soutenue par l'Europe depuis plus de dix ans, l'intelligence ambiante a pour objectif de remettre l'Humain au centre du processus de développement technologique. L'intelligence ambiante vise en effet le développement d'un environnement électronique intelligent et ubiquitaire qui aide les utilisateurs dans leurs tâches quotidiennes et professionnelles en respectant cinq critères : l'ubiquité (services accessibles en tout lieu), la prise en compte du contexte (adaptabilité et pertinence), la réactivité (aux changements), la pro-activité (anticipation) et la non intrusion (respect de l'individu). Dans ce contexte, les utilisateurs deviennent de plus en plus nomades et exigent une plus grande liberté, tout en souhaitant conserver le maximum de simplicité.

L'intelligence ambiante et les objets communicants constituent ensemble un enjeu majeur pour les années à venir, tant sur le plan scientifique que du point de vue des applications industrielles, commerciales et sociétales.

2 Projet scientifique

Le LIP6 propose d'aborder ce thème à deux niveaux, l'un sur l'infrastructure de communication et l'autre sur des aspects liés directement à l'intelligence ambiante. Ce projet réunit treize équipes du laboratoire de quatre départements scientifiques. La spécificité du LIP6 dans le thème repose sur le fait d'avoir au sein du même établissement des forces sur toutes les couches d'une architecture informatique pour les systèmes intelligents dynamiques. Au travers des différents objectifs tracés dans la suite, l'ambition du laboratoire est de montrer que ces forces complémentaires peuvent donner lieu à des avancées majeures dans le domaine.

3 Défi “Données et interactions”

3.1 Objectifs

Ce défi concerne les aspects liés à l'infrastructure de communication dans le contexte des interactions établies par des éléments mobiles et communicants. Il est découpé en trois thèmes principaux :

- **Dynamique des interactions : Analyse et modélisation.** Le cœur de ce défi est d'analyser puis modéliser la dynamique entraînée par des équipements mobiles et tirer parti de ses propriétés afin d'améliorer le fonctionnement et l'utilisation de ces systèmes, ainsi que d'offrir un moyen de prédire la transmission des interactions au fil du temps.
- **Radio cognitive.** Composants portables capables d'analyser l'encombrement du spectre et d'établir une communication sans fil sur les bandes de fréquences disponibles avec les standards adéquats.
- **Données sur supports mobiles.** Dans cette thématique, nous proposons d'étudier des solutions distribuées pour le transfert, le stockage et l'interrogation de grandes masses de données partagées sur un grand nombre de nœuds mobiles qui forment un système distribué large échelle.

3.2 Verrous

- **Scientifiques.** Détection et caractérisation de la dynamique. Adaptation des algorithmes classiques pour le passage à l'échelle, la distribution et la limitation des ressources. La consommation et la surface de ces circuits non adaptées à des composants sans fil. Routage cognitif. Modèles de simulation intégrant couche physique et protocoles de communication. Réalisation intégrant la sémantique de l'application dans les protocoles de partage de données.
- **Techniques.** Logiciels pour la collecte de données. Association données/localisation. Stockage sur nœuds mobiles. Conception de circuits intégrés analogique et RF en technologies CMOS nanométrique. Utilisation de MEMS (MicroElectroMechanical Systems) pour améliorer les performances des circuits RF.
- **Sociétaux.** Exploitation de données personnelles. Internet des objets. Étude des réseaux sociaux. Politiques pour gérer l'accès dynamique au spectre et protection des droits de licence.

3.3 Positionnement national et international

Les thèmes abordés dans ce défi ont une place importante dans les principaux centres de recherches du monde, aussi bien académiques qu'industriels. Avec l'avènement des réseaux sociaux et l'introduction des objets communicants sans fil dans nos vies de tous les jours, la maîtrise de ces systèmes comme un tout est devenue la clé de voûte pour la proposition de solutions avancées qui soient à la fois efficaces et conformes aux exigences des utilisateurs. Sur les aspects "dynamiques des interactions", peu de groupes dans le monde mélangent des solutions issues des domaines des réseaux, de l'intelligence artificielle et de la théorie des graphes. Les contributions restent en général limitées à un domaine, comme par exemple les études des temps de contact et d'inter-contact en loi de puissance ou la détection de groupes communicants à l'aide de méthodes d'apprentissage à base de *clustering* dynamiques. En ce qui concerne les réseaux cognitifs, le LIP6 se place en position originale en mettant ensemble une équipe sur les réseaux et une autre équipe sur le matériel. Enfin, des travaux sur le partage de données entre mobiles ont conduit au développement de middleware spécialisés permettant uniquement le développement d'applications dédiées. Nous souhaitons plutôt développer une approche ouverte, en définissant un modèle de données commun proche de RSS ou ATOM.

3.4 Positionnement du projet par rapport aux défis généraux des départements.

Les défis listés ci-dessous s'intègrent parfaitement dans les objectifs du département. D'ailleurs, un projet LIP6 entre les équipes NPA et Complex Networks a pu explorer une partie des problèmes rencontrés lors de la modélisation des interactions entre nœuds mobiles. Ce projet a également permis d'identifier un certain nombre de problèmes ouverts dont ceux décrits dans ce document. Une autre projet (en cours) intra-département (RSR) porte sur la géolocalisation d'objets communicants du type RFID. Aussi, une thèse co-encadrée par les équipes PHARE et CIAN vient de démarrer sur la radio cognitive.

4 Défi "Intelligence ambiante"

4.1 Objectifs

La conception et la réalisation d'un système d'intelligence ambiante (SIAM) soulève de nombreux défis scientifiques, technologiques et sociétaux. Nous nous intéressons ici à des SIAM de grande ampleur (par exemple les villes intelligentes) où des éléments ou parties du système peuvent recouvrir aussi bien des entités physiques (capteurs, objets intelligents) que des humains (utilisateurs, décideurs, etc.). Nous nous focalisons sur quatre thèmes fortement corrélés :

- **Modélisation et Simulation Multi-Echelle.** L'ouverture des SIAM et leur hétérogénéité exige des modèles de natures et d'échelles différentes. La conception physique de tels systèmes est très coûteuse ce qui nécessite des modélisations et des simulations multi-échelles en vue de leur prototypage virtuel. Cela soulève deux grandes difficultés : i) la conception et la mise en œuvre de méthodologies permettant à différents modèles soit d'interagir directement (en simulation ou une fois déployés), soit de définir des processus de conversion de modèles qui font qu'un même processus, tâche, entité, pourra être modélisé de plusieurs manières (différentes échelles, niveaux, détails) mais de manière cohérente ; ii) la simulation

de ces systèmes tout en tenant compte des différentes constantes de temps des parties physiques et tirant parti des machines de simulation multi-coeurs disponibles (parallélisation et couplage efficace).

- **Algorithmique répartie et protocoles de coordination.** La modélisation, la conception de protocoles de coordination auto-organisant (prise en compte du contexte, autonomie et intelligence des entités, tolérance aux fautes et aux attaques), la vérification formelle et la validation expérimentale des SIAM requiert de nouveaux modèles de calcul et de preuve. Par conséquent de nouveaux algorithmes doivent être conçus afin de résoudre les problèmes spécifiques aux SIAM (le manque de vision globale, connectivité dynamique, partitionnement du réseau et besoin de caractériser des périodes de stabilité) pour la terminaison pour les protocoles de coordination.
- **Vérification et analyse comportementale en contexte.** Notre objectif est de proposer un environnement de vérification des SIAM, qui combine les points forts des approches de modélisation formelle des systèmes répartis et l'analyse dynamique des comportements. Nous proposons de faire réaliser la vérification par des agents car un système AMI est intrinsèquement décentralisé et en évolution constante. Les analyses comportementales serviront non seulement lors des phases de conception mais aussi dans un but applicatif pour améliorer la proactivité des agents. Il est important de noter que les outils actuels de vérification ne sont pas adaptés aux applications de type ubiquitaire. Leur adaptation nécessite une collaboration étroite avec des spécialistes en vérification. La validation expérimentale des solutions va demander à son tour une collaboration avec des spécialistes en réseau et robotique expérimentale.
- **Distribution de l'intelligence sur l'infrastructure.** L'infrastructure répartie est au cœur des SIAM. Les capteurs, les données et les processus sont répartis sur des systèmes hétérogènes et l'interprétation des données ne peut pas être centralisée. Un premier objectif de cet axe est de définir quels calculs et quelles opérations doivent être effectués sur chaque entité du système, et comment ces différentes entités peuvent inter-opérer. Nous envisageons une agentification du système, en termes d'observation des données, d'actions possibles sur l'environnement, de prise de décision sur ces actions, d'interaction avec les autres agents et, éventuellement, d'apprentissage et d'adaptation. Une deuxième question au cœur de cet axe est celle de la notion de capteur intelligent : quelle quantité et quelle nature de traitement sur les données doivent être faites au niveau capteur ; quelle granularité, en fonction de l'application (performances disponibles, modèles de traitement de données requis). Enfin, il est nécessaire de construire une représentation du contexte qui soit manipulable par les différentes entités du système. Cela nécessite de pouvoir intégrer des données souvent hétérogènes et de manière asynchrone.

4.2 Verrous.

- **Scientifiques.** Compromis entre économies de calcul et fidélité de la simulation ; Composabilité et interopérabilité des modèles de calcul ; Cyber-Physical Systems, Evaluation des systèmes multi-niveaux ; Modélisation de comportements émergents dans les SIAM, Conception de protocoles et la diffusion/recherche d'information et la coordination dans un monde hétérogène et ouvert ; Intégration de données dans les capteurs, Modèles contextuels ; Distribution des calculs (traitement de données) sur les différents supports matériel ; Performance ; Ressources énergétiques nécessaires (équipement mobiles, équipements autonomes) aux traitements de données.
- **Techniques.** Simulation multi-échelle multi-disciplines parallèle sur machine multi-coeurs ; Parallélisation de la simulation et de la vérification ; Déploiement d'une infrastructure de SIAM ; Différentes abstractions des entités, différents modèles de calcul, simulation multi-niveaux.
- **Sociétaux.** Continuum du virtuel au réel. L'ensemble des techniques de prototypage virtuel devrait permettre de passer de manière continue de la simulation à la réalisation physique opérationnelle. Aide à la personne ; Commerces, loisirs, réseaux sociaux ubiquitaires.

4.3 Positionnement national et international.

L'intelligence ambiante (AmI) constitue donc un enjeu majeur pour les années à venir. C'est un "hot topic" de recherche au niveau international. Il existe déjà beaucoup d'études ("smart rooms", "open cities", "smart cities") et quelques réalisations concrètes qui intègrent capteurs et systèmes intelligents (dont plusieurs à l'international et aussi en France – voir par exemple l'IRCICA à Lille <http://www.ircica.univ-lille1.fr/>). De nombreux colloques et conférences internationales sont également dédiés au sujet (voir par exemple

conferences.acadiau.ca/ant-12/). Au niveau national, l'AmI est aujourd'hui une des thématiques prospectives du CNRS. Deux écoles d'été nationales soutenues par le CNRS, l'INRIA, le GDR-I3, lui ont été dédiées (voir par exemple <http://www.polytech-lille.fr/ecole-d-ete-du-cnrs-intelligence-ambiante-art240.html>). On peut aussi citer de nombreux projets industriels et/ou territoriaux en partenariat avec des laboratoires de recherche qui sont partis des "smart rooms", pour aller de plus en plus vers des environnements ouverts : les "maisons intelligentes", les "bâtiments intelligents" (voir par exemple le projet ADREAM du LAAS-CNRS), les "smart cities", et aujourd'hui le projet international "smart(er) planet" d'IBM. Les questions soulevées par ces projets sont plus larges que l'AmI (par exemple, opendata, simulation) mais recourent très souvent les thématiques de l'AmI.

Le LIP6 regroupe des compétences dans plusieurs domaines de l'intelligence ambiante (matériel, réseau, logiciel, Intelligence Artificielle) et c'est probablement l'un des rares laboratoires en France disposant d'une telle couverture. En favorisant la synergie entre ces compétences, il peut apporter une contribution réelle et visible sur les thématiques de l'intelligence ambiante et leurs fondements théoriques.

4.4 Positionnement du projet par rapport aux défis généraux des départements.

L'intelligence ambiante a de nombreuses interactions potentielles avec les deux thématiques prospectives "nouveaux systèmes" et "e-santé". Dans le premier cas les synergies porteront sur les aspects large-échelle de l'intelligence ambiante et de l'analyse, la modélisation et la simulation multi-niveaux qui nécessiteraient des systèmes multi-cœurs par exemples. Dans le cas du thème "e-santé", de nombreuses interactions peuvent être envisagées allant de la modélisation de l'utilisateur jusqu'au déploiement d'agents intelligents sur dispositifs d'aide et d'assistance aux utilisateurs (voir par exemple l'appel à projet Européen récurrent sur l'*Ambient Assisted Living*). Au niveau des départements, les liens sont naturels puisque cette thématique regroupe des chercheurs de 4 départements (sur 5) du LIP6. Les départements SOC et DESIR (distribution de l'Intelligence sur l'infrastructure, la simulation et la modélisation multi-niveaux). MoVE, REGAL, NPA et DESIR (algorithmique répartie, interactions intelligentes et mobiles, vérification de comportements en contexte, etc.).

5 Analyse SWOT

Points forts : Complémentarités rares au sein du même laboratoire ; équipes bien établies au niveau mondial dans leurs domaines respectifs ; abordage intégrée de deux niveaux habituellement considérés de façon séparées (infrastructure et intelligence ambiante).

Points faibles : Manque d'une équipe "traitement du signal" ; manque d'un laboratoire pour effectuer des expérimentations *in vivo*.

Risques liés au contexte : Perte de l'équilibre entre théorie et pratique.

Possibilités liées au contexte : Applications industrielles importantes ; collaborations avec des collectivités locales ;

6 Stratégie et mise en œuvre

Coopération interne/externe : Thèses en co-encadrement (une vient d'être lancée sur la radio cognitive), montage de projets internes au laboratoire et avec d'autres institutions en France et en Europe.

Liens avec acteurs sociaux économiques et recherche :

Développement de compétences, qui inclut la formation : Transfert des compétences vers les masters gérés par le LIP6, proposition de projets aux étudiants sur les thèmes abordés, enseignement à distance et à l'étranger avec des universités partenaires.

Moyens existants et recherche de moyens : Les équipes concernées ont montré dans le passé une très forte activité de montage et de participation dans des projets à des différents niveaux : ANR, Europe, Région, Labex (Smart) et Equipex (FIT).



Vague D : Campagne d'évaluation 2012-2013

2.2 Projet Transverse 2014 – 2018 - Thème Fiabilité/Sécurité

1 Présentation du Thème : Fiabilité/Sécurité

1.1 Périmètre du thème

Les systèmes d'information constituent une part importante du fonctionnement des entreprises et il est nécessaire d'assurer un niveau de sécurité et de fiabilité qui garantisse la pérennité de ces activités. S'il est donc clair que c'est un enjeu majeur, les mots fiabilité et sécurité recouvrent des sens différents selon les communautés de recherche en informatique. Nous avons identifié les besoins suivants :

Fiabilité :

- Assurance de la qualité de service ;
- Confiance/Qualité des résultats produits par un logiciel (par exemple qualité numérique) ;
- Sûreté de fonctionnement - Garantie de résultat (qui doit correspondre à un cahier des charges) ;
- Tolérance aux fautes (pannes franches).

Sécurité :

- Protection des messages et des données (pour assurer confidentialité, authenticité et intégrité) ;
- Protection contre des attaques extérieures ;
- Détection des fuites et mécanismes pour les contrer ;
- Tolérance aux fautes (attaques).

1.2 Contexte

- Le passage à l'échelle (calcul sur grille, vieillissement des circuits, densification des composants, ...) augmente de façon importante les fautes et il devient impératif d'accroître la tolérance aux fautes.
- Les attaques par canaux cachés constituent un domaine de recherche assez récent et très actif. L'ennemi peut en effet tirer parti des spécificités physiques d'un composant (mesure de consommation électrique par exemple). Ces attaques sont souvent simples à mettre en œuvre et plus puissantes que les attaques classiques.
- La fiabilité des systèmes informatiques courants n'est pas satisfaisante (plantages, non disponibilité, etc).

La question qui se pose naturellement est donc : que peut-on faire pour obtenir des systèmes plus fiables et/ou présentant des garanties de sécurité ?

2 Projet scientifique

Nous proposons deux défis en lien avec les tendances fortes du domaine : le premier défi est de comprendre et de maîtriser les fuites en proposant des contre mesures, que ce soit au niveau matériel/logiciel ou à un niveau plus théorique. Le second défi se positionne autour de la fiabilité des logiciels avec pour objectif de pouvoir prévenir les défaillances d'un programme lors de sa conception, de sa programmation et de son exécution.

Forces et spécificité du LIP6 dans le thème. La force du LIP6 dans ce domaine réside dans la complémentarité de ses équipes : par exemple pour le premier défi on peut citer des équipes travaillant au niveau du matériel (contre mesures), au niveau de l'arithmétique et de l'algorithmique, mais aussi des équipes qui étudient de manière plus théorique comment modéliser et détecter les fuites.

3 Défi 1 : Comprendre et maîtriser les fuites

3.1 Objectif du défi (coordinateur : R. Chotin-Avot)

Le développement de l'internet fixe ou mobile confère une place importante aux systèmes d'information qui véhiculent au quotidien les données numériques échangées sur l'ensemble de la planète. L'enjeu de la sécurisation des données est donc au cœur de la société de l'information.

Il existe de nombreux algorithmes cryptographiques permettant d'assurer la protection des données (*i.e.* confidentialité, l'authenticité, etc ...). Ces algorithmes sont de plus en plus souvent embarqués dans les GSM (E0), cartes à puces, etc ... Le déploiement sur des architectures matérielles expose les primitives cryptographiques à des attaques *physiques* (mesure de la consommation pour retrouver des bits du secret, injection de fautes pour retrouver le secret, etc ...).

Même si les algorithmes cryptographiques reposent sur des problèmes algorithmiques réputés difficiles (factorisation, logarithme discret, résolution d'un système d'équations non-linéaire, ...), ces attaques physiques peuvent compromettre la sécurité des implantations des algorithmes cryptographiques.

L'objet du défi est d'aborder les questions d'attaques et de fuites d'informations sous l'angle à la fois matériel, logiciel et théorique.

Dans une première partie du projet, il s'agit d'analyser les fuites dans certains composants cryptographiques (par exemple lors des opérations arithmétiques). Les mesures (mesures électriques, rayonnement électromagnétique, température, etc ...) révèlent de l'information normalement cachée de manière explicite ou implicite (par exemple on peut lier la consommation électrique au poids de Hamming d'une valeur d'état interne). Dans un deuxième temps il faut proposer des mécanismes de protection (contre-mesures) pour s'opposer à ces attaques. La nature de ces contre-mesures est fortement liée à la puissance de l'attaque que l'on peut mettre en œuvre.

Une deuxième partie du projet consiste à combiner les attaques algébriques classiques et les attaques par canaux auxiliaires de la façon suivante : en utilisant un certain nombre de mesures on obtient des informations supplémentaires qui sont traduites par l'attaquant en un système d'équations non linéaires ; reste à résoudre ce système en utilisant (par exemple) les algorithmes développés dans PolSys pour retrouver la clé secrète du cryptosystème. De la sorte on possède une technique très puissante de cryptanalyse (souvent il suffit de posséder des mesures partielles). Cependant ces techniques amènent à des problèmes fondamentaux difficiles : par exemple, comme les mesures sont bruitées, il faut être en mesure de résoudre des systèmes non linéaires en présence d'erreurs.

Une dernière partie du projet est consacrée à des approches formelles pour la détection d'informations cachées et les mécanismes de protection, soit par la mesure de fuites d'information, soit par la détection d'attaques potentielles, qui sont spécifiées par des politiques de sécurité.

Précisons, que la notion de détection de fuite doit être comprise au sens large : on peut injecter des fautes ou attaquer un composant avec un rayon laser ce qui permet de modifier son comportement (de manière plus ou moins aléatoire). Là encore un attaquant peut en tirer avantage.

3.2 Verrous du défi

Les verrous scientifiques suivants ont été identifiés :

- Définir un modèle théorique (par exemple en modélisant un système par un automate, éventuellement probabiliste) pour la robustesse, au sens de la résistance aux fuites.
- Tenir du compte du bruit dans les mesures (résolution d'équations avec bruit) dans des modèles probabilistes.
- Proposer des algorithmes (par exemple opérateurs arithmétiques) minimisant les fuites.

3.3 Enjeux industriels

Les attaques par canaux s'appliquent à un grand nombre de composants. Parmi ces composants il faut noter la place particulière de la carte à puce. La carte à puce est un secteur où l'industrie française est très développée.

On peut citer Bull (acteur historique et premier brevet) et son utilisation dans les cartes bancaires réduisant le taux de fraude de 50% en quelques années. De nouvelles applications sont apparues avec notamment le téléphone mobile, la carte santé, la télévision à péage, le passeport électronique et le transport (Navigo,...) et des acteurs industriels importants comme STMicroelectronics, Gemalto, Sagem, Ingénico, Oberthur.

Ces industriels possèdent, pour la plupart, une forte expertise dans le domaine des attaques par canaux cachés.

3.4 Forces du LIP6 pour relever ce défi

Cinq équipes (impliquant trois départements) ont manifesté un intérêt pour ce défi. Les équipes *PEQUAN - ALSOC - CIAN - MOVE - POLSYS* ont chacune une expertise dans le domaine de la sécurité :

- L'équipe CIAN (R.Chotin-Avot, H.Mehrez) travaille sur l'implantation matérielle des algorithmes de cryptographie et les FPGAs sécurisés.
- L'équipe ALSOC (N. Drach-Temam, E. Encrenaz, K. Heydemann, F. Wajsburt) travaille sur l'audit d'applications de sécurité, la mise en place de contre-mesures matérielles et l'adaptation dynamique du niveau de contre-mesures.
- L'équipe POLSYS est pionnière dans l'utilisation de techniques algébriques en cryptanalyse (Cryptanalyse Algébrique) et s'intéresse depuis à une extension de ces techniques pour les attaques par canaux cachés (Algebraic Side Channel Attacks).
- L'équipe PEQUAN possède une expertise reconnue dans la conception d'opérateurs arithmétiques pour la Cryptographie.
- L'équipe MOVE apportera sa compétence dans le domaine de la modélisation à base d'automates (éventuellement probabilistes).

Les équipes POLSYS, PEQUAN et CIAN collaborent depuis maintenant un an pour mettre leurs compétences en commun autour du projet LIP6 CRYPT2HARD qui vise à proposer un environnement de conception pour le prototypage rapide des algorithmes cryptographiques vers des architectures matérielles dédiées. Le défi proposé ici est bien plus large et ambitieux. On cherche à proposer un modèle théorique des fuites, pour ensuite proposer du logiciel/ matériel incluant des contre-mesures appropriées.

Parmi les faiblesses du LIP6 il faut noter qu'il n'y a pas de banc d'attaques physiques dans le laboratoire (toutefois ces compétences sont disponibles à proximité par exemple au LTCI).

3.5 Transversalité du défi

Les aspects que nous souhaitons mettre en avant sont :

- Mesures/contre-mesures : logiciel / matériel (CIAN, ALSOC, MOVE)
- Conception innovante de circuits intégrant la sécurité (CIAN, ALSOC, PEQUAN)
- Attaques algébriques et canaux cachés (informations implicites) (CIAN, POLSYS, MOVE)

3.6 Positionnement national et international du défi

Notre particularité est de pouvoir attaquer le défi sur ces aspects théoriques (modélisation par automates et modèles probabilistes) et pratiques (au niveau des architectures logicielles et matérielles). C'est une caractéristique rare.

3.6.1 Au niveau des architectures matérielles De nombreuses équipes françaises s'intéressent aux attaques matérielles et méthodes de contre-mesures. On peut citer le LIRMM (Montpellier) qui travaille sur les aspects test et sécurité dans le sens où les techniques de test actuelles des circuits qui se basent sur l'observabilité des points mémorisants peuvent nuire aux secrets des informations. D'autres travaux sont menés au LIRMM et au Lab-STIC (Lorient) au niveau des systèmes sur puce pour protéger les transactions (bus, caches, protocoles de communication). Pour ce qui est des attaques physiques, le LTCI (Telecom ParisTech) et plus particulièrement l'équipe COMELEC a mis en place un banc d'attaque de circuits et travaille sur les attaques en consommation et magnétiques, ainsi que sur des méthodes de contre-mesures pour ce type d'attaques. Les laboratoires PRISME (UVSQ) et XLIM (Limoge) se concentrent plus particulièrement sur les cartes à puce et proposent des contre-mesures logicielles. Enfin on peut citer le laboratoire Hubert Curien (St Etienne) qui travaille sur les implantations matérielles d'algorithmes de cryptographie.

Sur le plan international, une équipe dynamique en matière de cryptographie est le groupe Crypto de l'Université Catholique de Louvain (Belgique) qui travaille à la fois sur les attaques et contre-mesures matérielles ainsi que sur les implantations matérielles liées à la cryptographie. On peut également citer le groupe KRYPTO de

l'IAIK TU Graz (Autriche) et de C. Paar (Bochum, Allemagne) spécialisés dans l'analyse de la sécurité des primitives cryptographiques.

3.6.2 Modélisation à base d'automates Suite aux premiers travaux initiés dans les années 70 par Bell et La Padula, Lampson, Denning, Volpano et Smith par exemple, la prolifération des moyens de communication a récemment provoqué une explosion de la recherche sur ces sujets, au niveau national et international.

Concernant les modèles et le développement d'outils pour la détection de flots illégaux, on peut citer par exemple en France :

- l'équipe SSIR de Supelec (Ludovic Mé)
- l'équipe SERES de Telecom Bretagne (Frédéric Cuppens),
- l'équipe SECSI du LSV (Hubert Comon-Lundt et Jean Goubault-Larrecq),

et à l'étranger :

- le projet DISCOLAB de l'Université Rutgers (Livi Iftode) aux Etats-Unis,
- le projet CASED de l'Université Darmstadt (Heiko Mantel).

D'autre part, certaines équipes consacrent des travaux à la modélisation et au développement de protocoles offrant des garanties de sécurité comme la signature électronique de contrats ou le vote électronique, qui sont des problèmes notoirement difficiles, par exemple

- les équipes ACAC et INSS de l'université Macquarie en Australie,
- le projet ANR PROSE, une coopération entre les laboratoires LSV, VERIMAG et les équipes INRIA Cascade et Cassis.

En outre, plusieurs agences gouvernementales ont des départements "Intelligence and Defense" dédiés à cette thématique.

3.7 Analyse SWOT du défi

Points forts : Notre particularité est de pouvoir attaquer le défi sur ses aspects théoriques (modélisation par automates) et pratiques (au niveau des architectures matérielles).

Points faibles : Pas de laboratoire pour réaliser des Mesures Physiques ou faire des expériences (attaque par laser) au LIP6.

Risques liés au contexte : Équipes dynamiques travaillant sur des sujets connexes en France/Europe.

Possibilités liées au contexte : Enjeux Industriels forts : l'Industrie française est très développée dans le domaine de la carte à puce. On peut aussi citer la proximité d'acteurs majeurs du domaine : ST Micro, Gemalto, Sagem, Ingénico, Oberthur.

4 Défi 2 : Logiciel Fiable

4.1 Objectif du défi (coordinateur : G. Muller)

De nos jours, des missions de plus en plus critiques sont confiées à des logiciels informatiques de plus en plus complexes. Cela pose la question de la fiabilité de ces logiciels. Cet enjeu est donc au centre des approches de conception et de développement. Elle trouve un écho tout particulier dans le domaine des systèmes répartis, qui est l'une des forces du LIP6.

Notre objectif est de proposer une approche intégrée pour la définition d'un langage de haut niveau, de son environnement d'exécution et de l'ensemble des outils d'analyse (formelle) et de compilation/génération de code et de tests permettant d'améliorer la fiabilité des programmes écrits dans ce langage. Plus précisément, nous voulons prévenir les défaillances d'un programme lors de sa conception, sa programmation et son exécution. Les défaillances traitées vont des comportements anormaux d'un programme, au crash d'exécution ainsi que les dénis de service.

4.2 Verrous du défi

Le verrou principal, scientifique et technique, est d'intégrer dans la chaîne complète de développement de programmes (conception, programmation, exécution, maintenance) les approches et outils actuellement cantonnés sur une phase de développement. Cela va de la modélisation d'un problème, au modèle de calcul du programme,

aux analyses statiques sur celui-ci, à l'abstraction du matériel par des machines virtuelles et des environnements d'exécution gérés, à des solutions de monitoring et de vérifications dynamiques. En particulier, il faut exploiter les contraintes non fonctionnelles ayant un impact sur la fiabilité.

En marge de ces aspects, s'exprime également le besoin pour de nouvelles méthodologies de développement permettant l'usage de ces outils dans le contexte de logiciels plus fiables. Un problème actuel est la place et l'utilisation de ces méthodes et outils dans le contexte de logiciels certifiés (par exemple, dans des domaines critiques).

D'un point de vue sociétal, cela se traduit par une plus grande confiance dans le logiciel utilisé dans la plupart des objets numériques du quotidien. Cela est important car des missions de plus en plus critiques sont confiées à ces systèmes, par exemple dans le domaine des communications, des transports, etc.

4.3 Forces du LIP6 pour relever ce défi

Quatre équipes ont manifesté un intérêt pour ce défi : APR, DECISION, MOVE et REGAL (réparties sur deux départements, RSR et DESIR). Les équipes issues de RSR sont spécialistes de différents aspects de la conception et de la réalisation de systèmes répartis. L'équipe issue de DESIR est spécialiste en modèles décisionnels.

Compétences de chacune des équipes :

- l'équipe APR (E. Chailloux, F. Peschanski) travaille sur les modèles de calcul, de concurrence, les systèmes de typage, l'abstraction du matériel, la couverture structurelle non intrusive et l'intégration de composants logiciels certifiés.
- l'équipe DECISION (C. Gonzales, P.-H. Wuillemin) travaille sur différentes formes d'inférences dans les modèles décisionnels, en particulier dans le cadre des approches probabilistes et multi-critères.
- l'équipe MOVE (S. Baair, B. Bérard, C. Dutheillet, L.-M. Hillah, F. Kordon, E. Paviot-Adet, D. Poitrenaud, N. Sznajder, Y. Thierry-Mieg) possède des compétences en modélisation, en théorie du contrôle, en exploration exhaustive des états d'un système (par *model checking*) et sur l'extraction de modèles depuis un programme.
- l'équipe REGAL (B. Folliot, J. Lawall, G. Muller, G. Thomas) travaille sur la vérification de règles de réécriture (logiciel Coccinelle), les machines virtuelles, l'analyse de traces et le refactoring de logiciels (également avec Coccinelle).

Les trois équipes couvrent la conception d'approches et d'outils pour prévenir les défaillances lors de la conception, la programmation et l'exécution de logiciels actuels et futurs. Nous le détaillons comme suit :

- **Conception** : Modélisation (techniques, nouveaux langages) ; Contrôlabilité d'une solution ou d'un problème ; Vérification formelle (*model checking*) ; Définition de propriétés (sûreté/accessibilité, causale/temporelles, quantitatives).
- **Programmation** : plus grande expressivité, généricité et sûreté des langages ; Outils de vérification (*software checking*, analyse statique, etc.).
- **Exécution** : exécution confinée dans des machines virtuelles (Managed Runtime Environments) ; Monitoring (modèle associé pour superviser l'exécution) ; Génération et exploitation des messages issus de traces d'exécution.
- **Tests** : génération automatique de tests pertinents à partir des spécifications du problème (fonctionnelles et non fonctionnelles), contrôle de l'optimalité des séquences de tests.

4.4 Approche et transversalité du défi

Le tableau ci-après ébauche quelques pistes de travail.

Approche	Technologies existantes
Conception de nouveaux langages	<ul style="list-style-type: none"> - Symmetric Nets with Bags - HironML (migration de threads) - SPOC (prog CPU/GPU) - pthreads/pigraphs (PI-calcul)
Association à des machines virtuelles d'exécution adaptées	<ul style="list-style-type: none"> - VMkit - ZAM sur PIC - ZAM dans navigateur Web
Analyse et monitoring	<ul style="list-style-type: none"> - Coccinelle - libDDD/libSDD, PNXDD, CPN-AMI - Evinrude - ZamCov (couverture MC/DC)
Génération et contrôle des tests	<ul style="list-style-type: none"> - aGrUM (logiciel d'inférence probabiliste) - BCC (Behavioral Consistency Checker) qui travaille sur des diagrammes de séquences UML

Nous identifions les actions de coopération suivantes :

- MOVE/REGAL : coupler les machines virtuelles à des solutions de monitoring (Evinrude) ou à un moteur de model checking.
- APR/REGAL : Implémentation de langages sur des machines virtuelles VM-KIT
- APR/MOVE/REGAL : Réalisation de la chaîne complète : implantation de langages sur des machines virtuelles posées sur des bibliothèques de model checking
- DECISION/MOVE : Génération automatique de séquences de tests comportementaux pertinents à partir de spécifications WSDL (web services). Cette activité démarre fin 2012 dans le cadre du projet FP7-MIDAS impliquant les équipes DECISION et MOVE.

4.5 Positionnement national et international du défi

L'originalité de ce défi relève de la conception intégrée de l'ensemble des outils relatifs à la conception et au développement de programmes. À notre connaissance, il existe de nombreux outils qui adressent chacun un ou plusieurs points du spectre.

- **Machines virtuelles et systèmes d'exploitation** : Micro-noyau certifié seL4 développé (NICTA), micro-noyau Singularity reposant sur une machine virtuelle modifiée .NET (Microsoft), machine virtuelle langage de recherche Maxine (Oracle).
- **Compilation et Analyse statique** : Compilation certifiée du langage C (Xavier Leroy, INRIA-Rocquencourt), Saturn (Stanford), CIL (Berkeley), KLEE - execution symbolique pour trouver des bugs (Standord, groupe de G. Candea à l'EPFL), analyseur statique de Coverity (produit issu de stanford).
- **outils et techniques de vérification** : outils développés en France à renoble (VERIMAG), Nantes (IRC-CyN), en région parisienne (LSV, LIPN, LIAFA), à Bordeaux (LaBRI). Il existe également des outils développés à l'étranger ; citons les plus célèbres à la NASA, à CMU, à Rice, à Turin, à Oxford, ou à Austin.

4.6 Analyse SWOT du défi

Points forts : Compétences dans les différentes phases du développement et utilisation d'un logiciel : design, programmation, exécution.

Points faibles : Manque d'expertise dans un domaine comme la sécurité (voir cependant le défi 1 de ce thème).

Risques liés au contexte : Complexité inhérente des problèmes liés aux logiciels. Passage à l'échelle pour avoir un véritable impact.

Possibilités liées au contexte : Le logiciel libre, qui permet d'observer le processus de développement des logiciels de taille réelle, et d'avoir un impact sur ce développement.

5 Stratégie et mise en œuvre

Coopération interne/externe : la coopération en interne pourra se faire au travers d'appel à projet localement ou nationalement (ANR) entre les équipes et départements du LIP6. Nous planifions de travailler avec les acteurs majeurs du domaine (comme par exemple le LTCI pour les mesures physiques ou le LIENS pour les aspects sécurité).

Liens avec acteurs sociaux économiques et recherche : Les participants de ces défis ont des contacts avec les principaux acteurs du domaine (Sagem, Thales, Oberthur, Gemalto, ...).

Développement de compétences, qui inclut la formation : Plusieurs membres du thème sont impliqués dans la création d'une spécialité au sein du Master d'Informatique de Paris 6 (Sécurité, Fiabilité et Performance Numérique).

Moyens existants et recherche de moyens : Nous nous appuyerons sur des projets institutionnels. A titre d'exemple, l'activité sur le test démarre fin 2012 dans le cadre du projet FP-7 MIDAS.



agence d'évaluation de la recherche
et de l'enseignement supérieur

Unités de recherche

Vague D : Campagne d'évaluation 2012-2013

2.3 Projet Transverse 2014 – 2018 - Thème LSDP

1 Présentation du thème "Traitement distribué de données distribuées"

D'après IDC (International Data Corporation) le volume de données dans le monde va être multiplié par 50 dans la prochaine décennie: 750 exabytes (750 milliards de gigabytes) d'information numérique ont été créés en 2006 et plus de 1000 en 2010. Si le défi des dernières décennies a concerné le stockage et l'accès aux données, les problèmes aujourd'hui sont plutôt de traiter les grands volumes de données et d'en extraire de l'information utile, synthétique et résumée. Les "big data" et les outils d'analyse associés font partie des grands axes de recherche mis en avant par le directeur de la recherche d'IBM (2011) et l'agence Gartner (2012).

De nombreux acteurs se sont déjà positionnés sur l'utilisation de calculs intensifs distribués à grande échelle pour faire face à la quantité de données à analyser. Ce type de traitement suppose que les données sont accessibles de manière centralisée. Toutefois, un volume important de données (documents personnels ou d'institution) ne le sont pas, ou pourraient le devenir. En effet, pour des raisons pratiques (taille) ou de confidentialité, il n'est pas possible ou souhaitable de déléguer les stockages de toutes les données générées par des activités personnelles et professionnelles à un seul organisme, d'autant plus lorsqu'il s'agit de croiser des données entre elles. Deuxièmement, la puissance des terminaux rend possible toute une série d'analyses au niveau local, qu'il ne serait pas économiquement intéressant de centraliser.

Ce thème a pour but de répondre au besoins actuels (données centralisées) et futurs (données distribuées) liés au traitement de grandes masses de données.

2 Projet scientifique

Nous distinguerons dans la suite deux défis, "orthogonaux", qu'il est nécessaire de relever pour pouvoir répondre aux exigences particulières de ce périmètre thématique. Le premier défi concerne le traitement, l'accès, et la fouille de données massives. Le second défi traite de techniques de traitements de masses de données où les données et les traitements sont distribués. Affronter simultanément les deux défis doit permettre de proposer des solutions pertinentes et efficaces au cadre visé, pour notamment aborder les divers cadres applicatifs et sociétaux que nous évoquons ci-dessous.

Équipes impliquées De nombreuses compétences nécessaires au développement de ce projet sont présentes au LIP6 dans différentes équipes: MALIRE (algorithmes d'apprentissage dans le cas de problèmes de grande taille), Bases de Données (pair à pair, services web, modèles de cohérence et de coût d'accès), NPA (transmission optimale de quantités importantes de données), Regal (systèmes répartis, réplication et cohérence) et Complex Networks (analyse de données modélisées par des graphes notamment dynamiques). Les synergies envisagées entre ces équipes sont les résumés dans la table 2.1.

Positionnement du projet par rapport aux défis généraux des départements Le passage à l'échelle des méthodes et le traitement de données massives, avec ou sans contraintes de distribution des données et des traitements, est une problématique clé de l'ensemble du département DAPA, c'est également une problématique centrale pour diverses composantes du département réseaux notamment pour les grands graphes, les réseaux de capteurs, etc.

TABLE 2.1 – Synergie entre les équipes du LIP6

BD → MALIRE	optimisation et distribution de traitements de données massives, échantillonnage, développement de techniques de plans distribués
MALIRE → BD	mise au point de modèles d'estimation des coûts de requêtes, apprentissage de paramètres pour l'optimisation de requêtes distribués.
CN ↔ MALIRE	analyse et la prédiction de la propagation de contenu dans les réseaux sociaux dynamiques, la prédiction de liens ou bien la recommandation sociale
BD → CN	indexation des données, stockage et optimisation de requêtes sur des graphes dynamiques
CN → BD	gestion de graphes dynamiques et algorithmiques associés.
NPA → BD	gestion de flux de données
Regal ↔ BD	problèmes de cohérence des données réparties

3 Défi Traitement de Données Massives

Objectif L'objectif du premier défi est le passage à l'échelle pour le traitement de grandes masses de données dans un cadre centralisé. Les problématiques principales concernent donc le *passage à l'échelle des modèles et des algorithmes existants* pour répondre à la croissance exponentielle des données à traiter. L'objectif principal est de développer des nouvelles solutions d'accès et d'analyse de données avec une complexité "sous-linéaire" afin de répondre efficacement à l'explosion des données mais également du nombre de traitements effectués. Les solutions proposées peuvent demander un compromis entre le coût et la qualité des résultats, utiliser des techniques approximatives (échantillonnage) et appliquer des solutions de factorisation et de parallélisation de traitements.

Verrous scientifiques et techniques Le passage à l'échelle des traitements sur des grandes masses de données requiert de s'intéresser plusieurs problématiques fondamentales, à savoir le développement de (i) nouvelles infrastructures de stockage et d'accès aux données car les principes classiques (ex. SQL) sont de plus en plus remplacées par des systèmes spécialisées (noSQL) pour différents types de données et applications, (ii) d'algorithmique efficace pour le large échelle : sous-linéarité, stratégies d'échantillonnage et compromis qualité de service/coût [2], (iii) de techniques de factorisation des traitements et de réutilisation de résultats partiels pour optimiser une collection de traitements similaires, et (iv) de la gestion de la dynamique.

Positionnement national et international En apprentissage, la thématique large échelle est mise en avant par de nombreuses équipes nationales et internationales : L. Bottou (Microsoft) sur le gradient stochastique, Y. Bengio (Université de Montréal) sur les architectures neuronales profondes, et Y. LeCun (New York University) sur la vision. En France, la problématique du traitement de données à large échelle est affichée comme un axe majeur de recherche (ex. LIG à Grenoble, l'IRIT à Toulouse, le LITIS à Rouen). Les services Web et le Web sémantique sont au centre des problématiques liées aux grands volumes en Bases de Données [1] étudiées par les équipes françaises (LIG Grenoble, LIRIS Lyon, INRIA Saclay, Montpellier) et au niveau international (InfoLab de Stanford aux USA, Max Planck en Allemagne, et LSIR à l'EPFL en Suisse). Il n'existe pas d'équipe qui s'intéresse de manière générale à des problèmes de passage à l'échelle d'algorithmes d'analyse et d'apprentissage sur des très grands volumes de données. Finalement, dans le domaine des réseaux, la plupart des équipes travaillant sur les graphes de terrain sont confrontées à des problèmes d'échelle, parmi lesquelles les plus proches du projet sont les université de Lyon (réseaux de senseurs), de Bordeaux (visualisation), et au niveau international, Stanford (réseaux sociaux ou issus de l'internet)

4 Défi Traitement Distribué de données distribuées et dynamiques

Objectif On considère un réseau (type internet) interconnectant des noeuds qui contiennent des données (sites Web, ordinateur personnel, téléphone portable). Dans de nombreuses applications de traitement de l'information, il est préférable, ou parfois imposé, que les données stockées dans un noeud du réseau y soient cantonnées, au moins partiellement. Le but de ce défi est de proposer des techniques de traitement distribué de données distribuées et dynamiques. La distribution des données et des calculs peut être réalisée à différents niveaux et à différents degrés pour obtenir des résultats optimaux tout en respectant les contraintes imposées

par le domaine applicatif. Il s'agit de trouver une solution décrivant où et quand effectuer quel traitement, qui apporte un équilibre entre centralisation et distribution tout respectant les niveaux de confidentialité et de coût (énergie/transport) souhaités. Les objectifs à moyen terme sont de constituer un groupe de travail qui permette de définir de manière plus précise les besoins nécessaires au développement du défi, ce qui nécessite le développement d'un ou plusieurs cadres applicatifs représentatifs des enjeux liés à ce défi.

Verrous scientifiques et Techniques Le premier verrou est lié au haut degré de synergie demandé : il sera nécessaire de développer de manière commune une solution qui permette un apprentissage distribué sur des données distribuées (équipe MALIRE), ce qui nécessite une maîtrise des données et plans d'exécution distribués (équipe BD), de gestion des flux d'information (équipe NPA) et de connaissance du graphe dynamique (équipe Complex Networks) des pairs. Finalement, avoir un contrôle du degré de confidentialité des données transmises est particulièrement important, ce qui est un problème non trivial avec des nœuds hétérogènes et non contrôlés.

Verrous Sociétaux Ce défi pose les problèmes (i) de la confiance des utilisateurs au niveau des algorithmes distribués, (ii) de l'atteinte d'une masse critique d'utilisateurs afin que le traitement soit d'une qualité suffisante, (iii) Du développement d'un réseau permettant de coordonner l'activité sur le réseau.

À notre connaissance, il n'y a peu d'équipes travaillant exactement sur la thématique du traitement distribué avec des données distribuées dans des réseaux P2P - les traitements ou les données peuvent être distribués mais pas en même temps. Au niveau apprentissage, des équipes travaillent sur le thème de "privacy preserving machine learning" mettant en jeu un grand nombre de nœuds du réseau, comme par exemple aux USA l'université de Californie [4] et l'université du Texas [3], et le L3S Research Center en Allemagne [6]. De nombreuses équipes travaillent sur le problème de la confidentialité des données, mais la plupart se concentrent sur les techniques de cryptographie. Les travaux les plus proches du défi sont ceux Dork chez Microsoft [5]. Une large partie des activités de recherches récentes en Bases de Données se concentrent sur la gestion de données distribuées sur le web. En France on peut citer les travaux des équipes LEO (I. Manolescu) et WebDam (S. Abiteboul) autour de la gestion de données XML et RDF, Zenith (P. Valduriez) sur l'intégration de données scientifiques et Hadas (C. Collet) sur le partage de données par coordination de services. Au niveau international on peut citer les travaux des équipes de J. Widom (Univ. de Stanford) sur le "Large Scale Graph Processing", et G. Weikum (Max-Planck Institut) sur la recherche d'information distribuée.

5 Analyse SWOT

Points forts : Complémentarité des compétences des équipes impliquées BD, CN, MALIRE, NPA. Originalité des thématiques du second défi.

Points faibles : Il existe une concurrence importante sur certaines thématiques du premier défi. Par ailleurs il faudra être capable de faire émerger une synergie forte entre les équipes du LIP6.

Risques liés au contexte : Le second défi aborde des problématiques nouvelles et originales sur lesquelles il est plus difficile de garantir des succès

Possibilités liées au contexte : La problématique "big data" est très actuelle et de nombreux acteurs académiques comme industriels sont engagés ou vont s'engager dans des travaux connexes. Cela multipliera les opportunités de collaborations académiques et de dépôt de projets collaboratifs.

6 Stratégie et mise en œuvre

Il existe d'ores et déjà des collaborations entre certaines équipes (MALIRE/BD, MALIRE/CN, etc) entamées à l'occasion de divers projets ou co-encadrements d'étudiants en thèse. Il faudra construire sur ces collaborations pour faire naître une synergie plus large autour des défis identifiés.

Par ailleurs, certaines thématiques font déjà l'objet de projets collaboratifs naissants ou à venir nationaux (ANR) et européens. Il paraît donc acté que les équipes concernées sont déjà potentiellement engagées et compétentes pour répondre aux défis.

Le contexte des big data étant très dynamique aujourd'hui il sera certainement aisé de trouver des collaborations au niveau national et international, académique et industriel, pour aborder ensemble ces problématiques difficile et ainsi limiter les risques.

Références

- [1] P. Bernstein, M. Brodie, S. Ceri, D. DeWitt, M. Franklin, H. Garcia-Molina, J. Gray, J. Held, J. Hellerstein, H. V. Jagadish, M. Lesk, D. Maier, J. Naughton, H. Pirahesh, M. Stonebraker, and J. Ullman. The asilomar report on database research. *SIGMOD Rec.*, 27(4):74–80, Dec. 1998.
- [2] L. Bottou and O. Bousquet. L'apprentissage statistique à grande échelle. *Revue Modulad*, 42:61–73, 2010.
- [3] J. Brickell and V. Shmatikov. Privacy-preserving classifier learning. In *Financial Cryptography*, pages 128–147, 2009.
- [4] K. Chaudhuri and C. Monteleoni. Privacy-preserving logistic regression. In *NIPS*, pages 289–296, 2008.
- [5] C. Dwork. A firm foundation for private data analysis. *Commun. ACM*, 54(1):86–95, 2011.
- [6] S. Zerr and W. Nejdl. Privacy preserving document indexing infrastructure for a distributed environment. *PVLDB*, 1(2):1638–1643, 2008.



Vague D : Campagne d'évaluation 2012-2013

2.4 Projet Transverse 2014 – 2018 - Thème Optimisation

1 Présentation du Thème “Nouvelles approches en optimisation”

Ce thème vise à rassembler des compétences présentes au LIP6 en optimisation pour s'attaquer aux problématiques modernes de ce domaine. L'évolution des technologies de l'information et de la communication pose en effet des questions nouvelles qui sont l'occasion de développer des approches originales. Un spectre assez large d'activités de recherche liées à l'optimisation est présent au LIP6, comportant d'une part des aspects fondamentaux de la modélisation ou de la résolution de problèmes et d'autre part des applications originales dans différents secteurs de l'informatique. Des discussions menées au sein du laboratoire ont permis de faire l'état des forces présentes et de révéler des préoccupations scientifiques communes et parfois complémentaires des uns et des autres. Elles ont fait apparaître des liens et des défis susceptibles de favoriser l'échange et la confrontation de méthodologies et de problématiques. Ce document cherche à décrire ces préoccupations méthodologiques, souvent inter-équipes, qui témoignent du potentiel d'actions transverses autour du thème optimisation.

2 Projet scientifique

Le thème s'articule autour de deux défis. Le premier d'entre eux s'intitule “l'optimisation sur des instances modernes” et concerne les problèmes d'optimisation impliquant des instances aux caractéristiques particulières les rendant plus difficiles. Cela inclut par exemple des instances de très grande taille, qui ne peuvent être stockées intégralement en mémoire, ou des instances qui changent au cours du temps, ou qui sont révélées au fur et à mesure. Le second défi s'intitule “l'optimisation pour des objectifs modernes” et concerne des problèmes d'optimisation impliquant des critères particuliers, motivés par la nature complexe des objectifs poursuivis (objectifs multiples, équité, robustesse, précision, résolution distribuée). En outre, l'utilisation d'une approche algébrique dans la modélisation du problème pose de nouveaux problèmes algorithmiques et d'analyse.

Une spécificité du LIP6 est de réunir des compétences très variées autour du thème de l'optimisation qui rassemble 9 équipes complémentaires du LIP6 (APR, COMPLEX NETWORK, DECISION, PEQUAN, PHARE, RO, PolSys, SMA) issues de trois départements (CALSCI, DESIR, RSR). Les travaux actuels menés au LIP6 en optimisation sont soutenus par de nombreux financements, en particulier les contrats ANR COCA, EXACTA, GeoLMI, Guepard, NeTOC, TaMaDi, TODO.

3 Défi “Optimisation sur des instances modernes”

Optimisation avec contrainte d'espace

Quand les entrées d'un calcul d'optimisation sont constituées de très grandes données, par exemple les graphes des réseaux sociaux, ou le log d'un serveur, la contrainte d'espace de nos machines pose de nouvelles problématiques, en particulier dans les domaines suivants :

- *Le codage compact d'instances* afin de représenter les données respectant la contrainte d'espace.
- *L'optimisation manipulant des représentations compactes* qui serait fondamentalement différente d'une approche décodage/calcul/codage.

- *L’algorithmique pour flux ou streaming algorithms* qui ne s’autorise qu’un nombre restreint de passes (souvent une seule) sur les données.

L’objectif est alors de développer une nouvelle algorithmique pour ces contraintes d’espace, qui reste efficace sur ces grandes données réelles. Les problèmes d’optimisation qui se posent naturellement dans les graphes de terrain tels que la topologie de l’internet (routeurs et câbles entre eux), les graphes du web (pages web et liens hypertextes entre elles), les réseaux d’échanges divers (pair-à-pair, e-mail, réseaux sociaux), devraient fournir un cadre applicatif naturel à ces problématiques. Ces applications suggèrent également de s’intéresser à une autre source de difficulté liée à l’évolution permanente du réseau/graphes dans lequel l’optimisation s’inscrit.

Optimisation en contexte dynamique

Le contexte dynamique de certaines informations est à l’origine de problématiques intéressantes en optimisation, en particulier :

- *la ré-optimisation à partir des solutions passées*, afin d’adapter une solution aux changements incrémentaux des contraintes.
- *l’optimisation en ligne ou online algorithms*, où il s’agit de prendre une décision au fur et à mesure de l’arrivée des données. Un exemple est le calcul d’ordonnancement au fur et à mesure que des nouvelles tâches apparaissent. Un défi consiste à développer pour l’ordonnancement en ligne l’approche primale-duale qui est une technique unificatrice pour de nombreux problèmes en ligne.
- *la compilation de problèmes d’optimisation ou de décision* où il s’agit, lorsqu’une instance est révélée en plusieurs temps, de savoir comment résumer au mieux l’information utile pour calculer la solution optimale efficacement quand l’instance sera totalement révélée.
- *l’optimisation séquentielle dans l’incertain* où il s’agit de rechercher des stratégies (séquences de décision) optimales pour parvenir à un but dans un environnement incertain.

4 Défi “optimisation pour des objectifs modernes”

Optimisation en présence d’objectifs multiples

Le but est de développer des modèles et des algorithmes visant à prendre en compte plusieurs points de vue durant l’optimisation. Ces points de vues peuvent refléter les évaluations de différents agents ou joueurs dans un problème d’optimisation collectif, ou différents critères à optimiser simultanément. On souhaite s’intéresser en particulier aux thèmes suivants :

- *optimisation équitable multi-agents*. Il s’agit de considérer des problèmes d’optimisation (typiquement un problème d’allocation de ressources) dans lesquels la qualité d’une solution est appréciée par divers agents et mesurée par une fonction d’utilité individuelle. On s’intéresse alors à la recherche de solutions qui maximisent les utilités individuelles tout en répartissant le plus équitablement possible la satisfaction entre les agents ce qui pose des problèmes d’optimisation non-linéaires. La résolution de ces problèmes peut s’envisager de manière centralisée mais aussi en mode distribué et coopératif ce qui constitue un défi supplémentaire car chaque agent n’a alors qu’une vue partielle du problème.
- *aspects stratégiques et théorie des jeux en optimisation* (manipulation, algorithmes à vérité garantie). Dans bien des processus d’optimisation (vote, allocation de ressources, enchères combinatoires) les agents peuvent vouloir manipuler le système en communiquant des informations erronées pour obtenir des solutions qui leur sont plus favorables. En théorie des jeux algorithmique, on s’intéresse alors à concevoir des mécanismes à véracité garantie, dont le principe d’allocation n’incite pas les individus à mentir sur les données préférentielles qu’ils communiquent.
- *optimisation multi-critère combinatoire*. Il s’agit de problèmes d’optimisation combinatoire dans lesquels la qualité d’une solution doit être évaluée selon plusieurs points de vues non nécessairement réductibles à un seul. Les préoccupations actuelles portent sur l’approximation de l’ensemble des solutions Pareto optimales avec garantie de performances ou la détermination efficace de compromis spécifiques dans cet ensemble.
- *applications en réseau et système*. On s’intéresse en particulier au routage et à l’allocation de ressources, problèmes pour lesquels naturellement plusieurs objectifs sont à optimiser (comme le débit et la latence par exemple ou la qualité de service et la quantité de ressources mises en œuvre) et pour lesquels plusieurs agents exercent des demandes concurrentes sur une ressource.

Certification et garanties de performances pour l'optimisation

Traditionnellement la recherche de garanties de performances en optimisation vise à développer des techniques algorithmiques qui produisent des solutions approchées munis de garanties de ratio d'approximation. Il peut s'agir d'heuristiques gloutonnes, mais aussi d'approches primales-duales couplées à une analyse polyédrale.

Pour bon nombre d'applications (par exemple le calcul de trajectoires de particules dans un accélérateur, optimisation de trajectoires de satellites), la puissance de modélisation offerte par la programmation linéaire est insuffisante. Il est alors pertinent de modéliser ces problèmes d'optimisation par des fonctions algébriques ou transcendentes (bien souvent multi-variées). On est ici confronté à des problèmes de nature exponentielle en le nombre de variables, parfois même indécidable dans le contexte transcendant. A ces verrous, s'ajoutent des problèmes de fiabilité et de certification des résultats qui sont encore plus aigus dès que l'on sort du monde linéaire. De plus, le contexte bruité pose alors la question du sens à donner aux notions de robustesse/correction/exactitude. Les axes de travail envisagés sont les suivants :

- *Optimisation approchée* en général avec garantie de performance. Pour espérer une résolution efficace d'un problème, on doit souvent se contenter de solutions approchées, qui peuvent être le fruit d'une relaxation d'un programme à variables entières. Un compromis intéressant entre efficacité et optimalité est obtenu par des algorithmes *anytime*, qui sont interruptibles et dont la qualité de la solution augmente avec le temps d'exécution.
- *Optimisation algébrique* de fonctions polynomiales multi-variées sous des contraintes polynomiales. La difficulté dans ce contexte est la complexité exponentielle de résolution exacte de ces problèmes. L'objectif est d'améliorer la complexité de résolution en exploitant les structures spécifiques des problèmes rencontrés.
- *Transcendance*. Nouvelles méthodes pour l'optimisation de la précision des calculs en arithmétique des ordinateurs (i.e. combien de bits à stocker et à calculer, *precision* en anglais) par rapport à la qualité en précision finale voulue (i.e. combien de bits "corrects" à la fin, (*accuracy* en anglais). Souvent, ces problèmes sont encore résolus avec des heuristiques incomplètes ou une recherche exhaustive du fait du caractère parfois transcendant des fonctions considérées.
- *Robustesse et fiabilité*. La nature non-linéaire des problèmes étudiés pose des problèmes de robustesse et de fiabilité aigus surtout dans une modélisation basée sur diverses approximations. Un objectif est alors de définir des notions de conditionnement et de régularisation permettant de gérer correctement ces situations pathologiques.
- *Technique Hybride de calcul*. Le développement de techniques hybrides symbolique/numérique est un domaine émergent et prometteur de l'informatique du calcul scientifique dont le but est d'allier efficacité et fiabilité. Ces nouveaux paradigmes de résolution sont mis en place dans des contextes duaux : soit on utilise du calcul formel pour fiabiliser et améliorer des calculs numériques, soit on utilise des algorithmes de résolution numérique pour accélérer des procédés de résolution exacte. Un enjeu est bien entendu la mise en place de ces techniques pour l'obtention de certificats de positivité dans le contexte de l'optimisation globale.

5 Analyse SWOT

Points forts : Les points forts résident dans la concentration particulière de chercheurs issus des différents domaines liés à l'optimisation au LIP6 et leur complémentarité.

Points faibles : Les interactions entre les équipes existent mais doivent probablement être davantage développées, notamment entre les équipes appartenant à des départements différents voire à des communautés différentes. On pense par exemple à des synergies possibles entre les problématiques de routages ou d'allocation dans les réseaux, l'optimisation multi-agents/multicritère et la théorie des jeux algorithmique, ou encore au développement d'outils pour l'analyse de ratio d'approximation utilisables dans les différents sous-thèmes .

Risques liés au contexte : Malgré la complexité croissante des problèmes d'optimisation à résoudre dans l'industrie, il y a un risque d'évolution de la recherche industrielle vers des thématiques moins fondamentales et moins ambitieuses qui pourraient se contenter de résolution heuristiques de ces problèmes sans garantie de performance.

Possibilités liées au contexte : On dispose en Ile de France d'une concentration importante d'acteurs privés et universitaires travaillant sur les problèmes d'optimisation complexes. Les problématiques posées devraient rencontrer les préoccupations de partenaires industriels et académiques, en particulier sur des sujets tels que l'optimisation des réseaux de transport, des réseaux de communication, et le développements d'outils et de services pour l'internet. Réciproquement ces problèmes réels peuvent fournir des opportunités de valoriser les outils mathématiques et algorithmiques développés au laboratoire.

6 Stratégie et mise en œuvre

Coopération interne/ externe : une coopération interne est déjà en place entre différentes équipes au travers de plusieurs projets LIP6 mais aussi de quelques projets industriels, par exemple le projet Dem@Factory (optimisation de chaînes de numérisation de documents) qui implique trois équipes du LIP6.

Liens avec acteurs sociaux économiques et recherche : les chercheurs du LIP6 concernés par ce thème ont des liens avec de nombreux laboratoires du domaine, comme l'atteste le nombre de contrats ANR, les accords bilatéraux, et les projets INRIA qui impliquent nos équipes.

Développement de compétences, qui inclut la formation : les équipes du LIP6 concernées par ce thème transverse sont largement impliquées dans les spécialités de master informatique de l'UPMC, en particulier en algorithmique avancée, recherche opérationnelle, décision, calcul scientifique, graphes de terrains, optimisation dans les réseaux, intelligence artificielle.

Moyens existants et recherche de moyens : outre l'implication actuelle des équipes dans différents projets ANR relevant de l'optimisation, cette thématique transverse autour des nouveaux problèmes d'optimisation s'inscrit naturellement dans les axes de recherche du *GDR Recherche Opérationnelle*, du *GDR International Algorithmic Decision Theory*, et du *programme Gaspard Monge* récemment lancé par EDF et la fondation Jacques Hadamard (mécénat) *pour l'optimisation, la recherche opérationnelle et la gestion de l'énergie*, qui peuvent être des sources partielles de financement.



agence d'évaluation de la recherche
et de l'enseignement supérieur

Unités de recherche

Vague D : Campagne d'évaluation 2012-2013

2.5 Projet Transverse 2014 – 2018 - Thème NSR

1 Présentation du Thème "Nouveaux Systèmes Repartis"

Nous constatons depuis cinq ans deux évolutions majeures des infrastructures réparties. Pour les environnements à grande échelle, on assiste à une convergence vers des infrastructures virtuelles communes pour optimiser le partage de ressources (on parle d'informatique dématérialisée ou de *cloud computing*). Il s'agit d'échanger des données et d'intégrer des services sur une infrastructure physique hautement dynamique et massivement distribuée. Dans ces nouvelles infrastructures, les principaux défis sont liés au contrôle de l'utilisation des ressources dans le but d'assurer une continuité de service.

Pour les ordinateurs individuels, les architectures *manycores* visent à transformer l'accroissement du nombre de transistors en une augmentation de performance sans faire évoluer la fréquence afin de limiter la consommation des machines. Les machines à 4 cœurs sont d'ores et déjà la norme, on projette des machines personnelles à 100 à 200 cœurs d'ici 2020. Or plusieurs études récentes ont montré que les supports systèmes (les machines virtuelle et les systèmes d'exploitation) s'adaptent mal à la multiplication des cœurs et que les architectures matérielles actuelles ont des problèmes de passage à l'échelle. Les principaux défis sont liés aux accès aux données partagées et à la contention des bus mémoires.

2 Projet scientifique

Nous proposons d'aborder les défis liés à ces deux tendances fortes en analysant, modélisant et concevant des nouveaux protocoles, architecture et services pour les grands environnements dynamiques et les architecture manycores. Les défis concernent (1) la distribution massive en nous focalisant sur la dynamique et le passage à l'échelle, et (2) les architectures manycores en nous concentrant sur la gestion des données.

3 Distribution massive sur les grands environnements dynamiques

Objectifs du défi à moyen terme :

Les grandes infrastructures où un ensemble d'entités interagissent sont maintenant la norme. Il y a une grande diversité de ces environnements que ce soient des ensembles de machines hétérogènes (ordinateurs personnels, set-top box, périphériques mobiles, etc.) partageant des ressources et des services ou des nuages informatiques où les ressources physiques disponibles des grands centres de traitements sont offerts aux applications. Malgré leur diversité, ces environnements ont des propriétés communes qui soulèvent des nouveaux défis de recherche. Ils sont :

- Hautement dynamiques : à tout moment de nouvelles entités peuvent être ajoutées et retirées. De plus, la topologie du graphe de communication entre les entités est dynamique. Les liaisons peuvent être intermittentes et les entités (périphériques mobiles ou machines virtuelles dans les clouds) peuvent se déplacer.
- Fragiles : la multiplication des machines augmente la probabilité de pannes. La panne qui était considérée comme un événement rare dans les systèmes distribués traditionnels devient la norme.

De part ces propriétés, ces infrastructures deviennent de moins en moins contrôlables. Il n'est pas possible d'avoir une vision de l'état global. Les défis que nous abordons consistent donc (1) à analyser qualitativement ces environnements afin d'en extraire les propriétés clés, (2) proposer des protocoles et méthodes de conception pour offrir des services garantissant une certaine qualité malgré la vision partielle et incertaine du système et (3) offrir des méthodes pour dimensionner et prédire les évolutions des environnements.

3.0.1 Transversalité. Cinq équipes (ALSOC, DECISION MOVE, NPA, REGAL) issues de quatre futurs départements du LIP6 (DESIR, SOC, SysComp, NetSys) possèdent les compétences et outils permettant d'aborder ce défi. Les nouveaux protocoles et algorithmes proposés par les équipes NPA et REGAL seront vérifiés par les techniques de *model checking* proposées par MOVE. De même, une collaboration entre les équipes MOVE et ALSOC permettra l'utilisation et la conception d'outils (théoriques puis logiciels) pour capturer des notions de robustesse. Les équipes DECISION et REGAL collaboreront autour de l'optimisation d'algorithmes répartis prenant en compte l'incertitude. REGAL et NPA continueront leur collaboration sur des problèmes de dimensionnement et de modélisation stochastique de grands systèmes.

Verrous du défi

Analyse qualitative. Nous proposons des méthodes pour analyser les protocoles sur les grands systèmes. Il s'agit d'extraire des propriétés clés pour identifier par exemple les défauts des protocoles et conduire à leur amélioration. Au niveau de la vérification qualitative, les verrous sont de combattre l'explosion combinatoire inhérentes aux approches à base de *model checking*. Les facteurs de cette complexité sont, outre la taille, la prise en compte d'aspects dynamiques rendant un grand nombre de techniques de modélisation vulnérables à l'explosion combinatoire.

Analyse quantitative. Il est fondamental d'analyser et prédire finement les évolutions de grands systèmes. Le défi consiste à définir des modèles stochastiques prenant en particulièrement en compte la dynamique des systèmes : l'arrivée et le départ massif de nœuds, l'évolution du nombre d'entité et les défaillances.

Conception et algorithme. L'algorithmique répartie aborde depuis plusieurs années les problèmes de passage à l'échelle. La réflexion s'est essentiellement faite autour des modèles temporels (i.e., les hypothèses sur les délais de transmissions de messages). En revanche les aspects liés à l'extensibilité ont été moins explorés. La principale difficulté est liée à la forte volatilité des nœuds qui interdit une vision globale du système et implique des prises de décisions en fonction de visions très partielles voir fausses de l'environnement. Il faut donc définir une algorithmique tolérant les incertitudes sur les participants et capable de s'adapter aux changements dynamiques de l'environnement. Deux voies nous semblent prometteuses et adaptées aux grands environnements : l'algorithmique indulgente où l'incertitude est un paramètre d'entrée et les algorithmes probabilistes (par opposition à l'algorithmique répartie traditionnelle qui est déterministe). L'objectif final est de développer une algorithmique "agile" (en absence de contrôle et avec des décisions locales) qui prenne en compte de contraintes de qualités de services.

Positionnement national et international du défi

Les clouds sont un exemple de système massivement réparti et dynamique. Au niveau international, les grandes équipes de recherches en systèmes répartis (Berkeley, Cornell, Microsoft, Google, North Carolina) ont des programmes de recherches autour des clouds. Ces recherches se concentrent essentiellement sur la proposition de modèles de programmation pour les clouds, l'allocation dynamique de ressources pour satisfaire les contraintes de qualité de services, le stockage efficace de données, ou encore la fédération de plusieurs clouds (appelé Sky computing).

En France, en dehors du LIP6, nous avons identifié 7 équipes qui ont des activités soutenues autour des clouds : au LIG (équipes INRIA Sardes et Vasy), à l'IRISA (équipes INRIA ASAP, Myriad et KerData), à l'EMN et au LIP (Equipe Avalon) et à Sophia-Antipolis (équipe INRIA Oasis). Ces activités sont autour de la gestion de l'économie d'énergie, de la réservation de ressources autonomiques, de la réplication de données massive ou encore des techniques de migration dans le cadre du Sky computing. Le LIP6 collabore avec une grande partie avec ces équipes. Le laboratoire est d'ores et déjà reconnu comme un acteur important dans la

communauté Cloud Française de part sa participation au projet ANR MyCloud (REGAL) et Nu@ge (PHARE, REGAL). Malgré cette forte activité de recherche autour des clouds, il est y a très peu d'équipes qui s'intéressent au fondement algorithmique et à l'analyse de propriétés pour les clouds. En ce sens le positionnement du défi est original.

Sur l'analyse qualitative, il y a deux voies pour la vérification. La première concerne les approches "algébriques" (basées sur la preuve avec des outils/méthodes comme B, Coq, Z, Isabelle, etc.). Les principales équipes sont à Grenoble (Verimag, TIMA), l'X, dans les centres INRIA de Paris-Rocquencourt et Sophia-Antipolis. La seconde voie concerne avec les approches qualitatives et quantitatives plus centrées sur le model checking ou le contrôle/synthèse. Nous trouvons les acteurs de MeFoSyLoMa (le groupe dont le LIP6 est l'un des fondateurs et qui regroupe les équipes de vérification de la région parisienne): LSV, LIPN, LIP6, LACL, IBISC, CEDRIC. MOVE et ALSOC travaillent avec ces équipes dans ou hors cadre MeFoSyLoMa. En province, on retrouve l'Ircyn à Nantes, le LaBRI à Bordeaux, le LIFO à Orléans, le LAAS et Verimag à Grenoble qui développent des approches de type model checking

À l'international, il y a plusieurs pôles forts en model checking. Autour de la vérification stochastique, on retrouve l'Université de Turin (avec qui MOVE a des collaborations) et Oxford. Sur les réseaux de Petri il a une communauté importante en Allemagne. Enfin, sur les aspects temporisés, il y a la communauté UPPAAL centrée sur la Scandinavie. Il y a peu d'acteurs majeurs en Asie et quelques grands acteurs aux SA: CMU (Clarke, prix Turing 2007), Emerson (Texas), la NASA (Holzman avec SPIN/Promela), Microsoft-research et l'Université de Rice pour les centres principaux. Notre particularité et de nous concentrer sur la vérification de systèmes répartis asynchrones et présentant des aspects symétriques pour lutter contre l'explosion combinatoire.

4 Multi-cœurs généralistes

Objectif du défi à moyen terme :

Ces dernières années ont vu l'apparition d'architectures multicœurs dans les routeurs, les fermes de serveurs et dans l'informatique grand public (ordinateur de bureau, tablettes PC et même téléphones portables). Ces architectures comportent jusqu'à une cinquantaine de cœurs, et il est probable que dans les vingt prochaines années les ordinateurs portables d'entrée de gamme en embarqueront une centaine. Afin d'éviter le goulot d'étranglement du bus de communication unique, les cœurs sont désormais connectés en topologies relativement complexes, par exemple un ensemble de nœuds (grappes d'une dizaine de cœurs autour d'un bus) inter-connectés par un réseau maillé. La mémoire reste logiquement partagée, c'est-à-dire que n'importe quel cœur peut adresser n'importe quelle adresse ; néanmoins cette topologie entraîne un partitionnement physique entre nœuds. On a alors une mémoire à accès non-uniforme (NUMA) : contrairement à l'hypothèse généralement admise aujourd'hui, les temps d'accès mémoire dépendent fortement des localisations relatives du nœud de calcul et du nœud qui gère la mémoire adressée. Des mesures que nous avons effectuées au LIP6 sur un multicœurs Opteron Magny-Cours à 48 cœurs montrent que les temps d'accès à un emplacement mémoire peuvent varier de 1 à 400.

Cela modifie profondément le comportement des applications. Une conséquence immédiate est que la complexité des algorithmes, basée jusqu'ici sur l'hypothèse de la mémoire à latence d'accès uniforme, est profondément remise en cause. Il faut maintenant être capable de placer les traitements à proximité des données qu'ils accèdent, de façon à uniformiser et réduire ces latences. Ce problème a été étudié précédemment dans le contexte du calcul parallèle hautes performances, mais peu dans un contexte plus générique, comme l'informatique personnelle ou bien les serveurs d'application. Dans ce cas, les calculs et les traitements sont irréguliers, et, plus globalement, les conditions d'exécution des applications non prévisibles (leur charge de travail, la mémoire qu'elles utilisent, la charge de la machine due aux autres applications).

Transversalité. Six équipes (ALSOC, APR, MOVE, NPA, PEQUAN, REGAL) de trois départements différents (CALSCI, RSR, SOC) du LIP6 possèdent des compétences et des outils permettant d'aborder ce problème : au niveau matériel (ALSOC), au niveau du système d'exploitation et de l'environnement d'exécution (ALSOC, REGAL, PEQUAN), au niveau langage et compilation (APR, REGAL et PEQUAN), et avec des outils de vérification (MOVE, ALSOC et APR). Notre objectif est de faire émerger une collaboration entre ces équipes, de façon à aborder globalement tous les aspects du problème posé par les machines NUMA, afin de proposer

des solutions cohérentes de bout en bout, impliquant à la fois l'architecture, l'environnement d'exécution et le langage.

Verrous du défi

Autour de ce problème posé par le placement mémoire, deux principaux verrous doivent être adressés dans les prochaines années par la communauté scientifique.

Verrou 1 : Le patrimonial. Le premier verrou concerne le code patrimonial, c'est à dire le code écrit alors que les multi-cœurs n'existaient pas encore. Ce code n'offre aucune abstraction adaptée à la gestion du placement mémoire. Le problème posé est donc de placer les traitements et les données, par une approche "boîte noire", grâce à une observation du comportement des applications à l'exécution. Résoudre ce problème demande des compétences :

- en architecture, afin d'étudier des sondes matérielles pertinentes,
- en système et algorithmique distribuée, afin de migrer dynamiquement code et/ou données pour s'adapter dynamiquement aux conditions d'exécution changeantes,
- en compilation, pour générer du code binaire qui tolère la migration mémoire et l'hétérogénéité des processeurs, et finalement,
- en analyse des grands graphes dynamiques, pour mieux comprendre les propriétés de l'exécution (comme la localité), de façon à piloter l'environnement d'exécution adaptatif.

Verrou 2 : Le futur. Les langages impératifs et à objets, majoritairement utilisés aujourd'hui, ne sont pas adéquats pour ces architectures à mémoire non uniforme. En effet, le grain de l'objet ou de la structure mémoire est trop fin pour placer les données et le code : un placement à ce niveau entraînerait inévitablement de nombreuses communications inter-cœurs lors de l'envoi de messages entre objets. De plus, les langages impératifs autorisent des accès mémoire et des effets de bord non contrôlables, ce qui rend difficile un placement mémoire optimal. Proposer des langages ou des compilateurs adaptés au NUMA demande des compétences :

- en langage et compilation pour définir de nouveaux paradigmes permettant d'extraire des informations sur la localité d'accès mémoire et sur les propriétés de concurrence des programmes,
- en système et algorithmique distribuée pour construire des environnements d'exécution efficaces,
- en architecture pour étudier des algorithmes de gestion de cache adaptés à ces nouveaux paradigmes,
- en vérification formelle pour prouver que les propriétés de concurrence des langages sont correctes.

Positionnement national et international du défi

Depuis trois ans, de très nombreux travaux adressent les problèmes posés par les architectures à mémoire non-uniforme. Pour ne citer que les travaux majeurs de ces trois dernières années, on peut pointer les travaux menés sur le matériel avec des algorithmes permettant d'améliorer automatiquement la localité des traitements [3], sur les systèmes d'exploitation pour augmenter la localité [2, 1] et sur les compilateurs et les langages pour mieux prendre en compte la localité, que ce soit en proposant des environnements d'exécution plus efficace [4] ou des algorithmes plus efficaces [5]. Chacun de ces travaux adresse une des couches logicielles ou matériel individuellement, mais rares sont les travaux qui proposent de faire collaborer le matériel, les environnements d'exécution, les compilateurs et les langages de concert pour améliorer les performances des applications. La force du projet que nous proposons est de proposer d'attaquer les problèmes de performance posés par les architectures à mémoire non-uniforme de façon globale en faisant collaborer ces différentes couches.

En France, outre les équipes qui travaillent sur l'informatique haute performance (HPC), nous avons identifié quatre équipes qui ont des activités importantes autour des multi-cœurs généralistes, c'est à dire les multi-cœurs dans le cadre des serveurs et de l'informatique personnel : l'équipe PARKAS (INRIA/EMN), l'équipe CAPS (IRISA), l'équipe SARDES (INRIA Grenoble) et l'équipe LSR (LaBRI). Ces activités se concentrent sur les architectures (CAPS), les langages (PARKAS) et les systèmes (LSR, SARDE). Le LIP6 collabore avec toutes ces équipes. Le LIP6 est d'ores et déjà reconnu comme un acteur majeur dans la communauté multi-cœur française de part l'organisation des journées Informatique Massivement Multiprocesseurs et Multi-cœurs, et de part sa participation active au montage en cours de l'Action d'Envergure Nationale (AEN) Inria autour

des multi-cœurs. L'une des spécificités du Lip6 dans ce paysage est qu'il regroupe en son sein des équipes travaillant sur les différents domaines liés aux architectures à mémoire non uniforme.

5 Analyse SWOT

Points forts : Le LIP6 a la chance de couvrir un nombre de thématiques importantes qui permet d'aborder des défis transverses particulièrement ambitieux. Peu de laboratoires peuvent aborder les deux défis identifiés dans cet axe qui s'appuient fortement sur les compétences complémentaires en algorithmique répartie, modélisation, vérification, langages, prise de décision et architectures. L'impact des résultats obtenus peut être importante.

Points faibles : Les défis abordés sont ambitieux. Le risque est donc élevé et la concurrence internationale importante.

Risques liés au contexte : Ce thème nécessite la collaboration étroite entre plusieurs équipes.

Possibilités liées au contexte : Depuis plusieurs années, le laboratoire et l'UPMC a mis en place des outils, les projets internes LIP6 et les projets Emergence, qui permettent de faire émerger des projets collaboratifs en rupture. La proposition de ce thème est le fruit de l'utilisation de ces outils.

6 Stratégie et mise en œuvre

Coopération interne/ externe : Ce projet s'appuie des collaborations étroites existantes entre les équipes du LIP6 qui se sont concrétisées depuis 5 ans par des projets internes LIP6 commun et un projet Emergence de l'UPMC. Le LIP6 participe déjà à trois projets collaboratifs (deux ANR dont une coordonnée par le LIP6 et un projet sur les investissements d'avenir) dont les thèmes de recherche abordent certains problèmes posés par ce projet transverse.

Liens avec acteurs sociaux économiques et recherche : Ce projet interagira étroitement avec le groupe travail MeFoSyLoMa qui regroupe les équipes de vérification de la région parisienne.

Développement de compétences, qui inclut la formation: Ce projet s'appuie sur les étudiants issus de trois spécialités de nouveau master Informatique: les spécialités Systèmes Réparties, Systèmes Embarqués et Sécurité.

Moyens existants et recherche de moyens: Cette thématique est au cœur d'un des 5 thèmes scientifiques du Labex SMART.

7 Références

- [1] A. Baumann, P. Barham, P.-É. Dagand, T. L. Harris, R. Isaacs, S. Peter, T. Roscoe, A. Schüpbach, and A. Singhanian. The multikernel: a new os architecture for scalable multicore systems. In *22nd ACM Symposium on Operating Systems Principles, (SOSP 2009)*, pages 29–44, 2009.
- [2] S. Boyd-Wickizer, A. T. Clements, Y. Mao, A. Pesterev, M. F. Kaashoek, R. Morris, and N. Zeldovich. An analysis of linux scalability to many cores. In *9th USENIX Symposium on Operating Systems Design and Implementation, (OSDI 2010)*, pages 1–16, 2010.
- [3] E. Ebrahimi, C. J. Lee, O. Mutlu, and Y. N. Patt. Prefetch-aware shared resource management for multi-core systems. In R. Iyer, Q. Yang, and A. González, editors, *38th International Symposium on Computer Architecture (ISCA 2011)*, pages 141–152. ACM, 2011.
- [4] L. Gidra, G. Thomas, J. Sopena, and M. Shapiro. Assessing the scalability of garbage collectors on many cores. *Operating Systems Review*, 45(3):15–19, 2011.
- [5] M. Kuperstein, M. T. Vechev, and E. Yahav. Partial-coherence abstractions for relaxed memory models. In *Proceedings of the 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation, (PLDI 2011)*, pages 187–198, 2011.



Vague D : Campagne d'évaluation 2012-2013

2.6 Projet Transverse 2014 – 2018 - Thème Santé

1 Présentation du thème e-santé

1.1 Caractérisation

Selon la définition de l'OMS (1946), la santé de l'individu s'étend de la médecine jusqu'au bien-être et la santé publique y ajoute une dimension organisationnelle. En 2010, le rapport Lasbordes a donné la définition suivante de la télésanté (ou e-santé) : *utilisation des outils de production, de transmission, de gestion et de partage d'informations numérisées au bénéfice des pratiques tant médicales que médico-sociales*. Ainsi la e-santé trouve naturellement sa place comme thème transverse du LIP6, laboratoire d'informatique de l'UPMC, Université scientifique et médicale.

1.2 Vision prospective de l'évolution scientifique du thème

La e-santé est mentionnée comme un axe fort par les prospectives SNRI, AVIESAN, FP7 e-health, TIC et Santé (3 Pôles de Compétitivité franciliens) et Ministère de l'Industrie. A l'horizon 2020, la santé personnalisée apparaît comme la tendance de fond.

Localement, le Labex SMART (Human / Machine / Human Interactions in the Digital Society), dans lequel le LIP6 est fortement impliqué, a consacré l'un de ses cinq programmes à « Human Autonomy and Health ».

1.3 Défis principaux

Les principaux défis identifiés sur le plan international sont : assistance et formation ; imagerie médicale ; données et connaissances (fouille, analyse, traitement et fusion) ; dispositifs médicaux (économie d'énergie, Body Area Network,...) ; sécurité, fiabilité et sûreté.

2 Projet scientifique

2.1 Défis abordés

Deux défis principaux sont abordés de manière transversale : l'interprétation des données de la santé et la e-santé individualisée.

2.2 Forces

7 équipes du LIP6, appartenant à 5 départements (CALSCI, DAPA, DESIR, NetSys et SOC), sont déjà actives dans le domaine de la santé depuis de nombreuses années : PEQUAN (D.Béréziat, S.Dubuisson), ACASA (J.-G. Ganascia), MALIRE (C. Marsala, M.Rifqi), DECISION (P. Brézillon), MOCAH (J.-M. Labat, A.Yessad), APR (A. Valibouze) et SYEL (S.Feruglio, P.Garda, A.Pinna). Trois autres équipes sont intéressées par ce thème (ALSOC, CIAN et POLSYS).

2.3 Compétences

Les compétences des équipes du LIP6 intéressées par la e-santé peuvent se regrouper en trois grands domaines : *modélisation utilisateur* : serious game, context aware, assistance utilisateur ; *interprétation des signaux et des données* : image médicale, du numérique au symbolique, apprentissage automatique, fusion d'information ; *systèmes embarqués pour la santé* : capteurs, architectures et algorithmes embarqués, sécurité.

2.4 Transversalité par rapport à la structure du laboratoire et synergies

Le thème de la e-santé est clairement transversal, il rassemble des personnes travaillant en imagerie du département CALSCI, deux équipes en apprentissage du département DAPA, deux équipes en décision et serious games du département DESIR, une équipe en algorithmique du département NetSys et une équipe en systèmes embarqués du département SOC. Des synergies importantes sont possibles, parce que les techniques maîtrisées par ces équipes sont complémentaires et que leur combinaison peut apporter une amélioration des fonctionnalités ou des performances des systèmes destinés à la e-santé.

2.5 Spécificités du LIP6 dans le thème

Une spécificité remarquable du LIP6 est qu'il couvre une très large part des domaines scientifiques nécessaires dans le domaine de la e-santé.

3 Défi Interpréter des données de la santé

3.1 Objectif du défi à moyen terme

L'objectif principal de ce défi est d'élaborer une chaîne d'interprétation allant des données jusqu'aux connaissances, dans le domaine de la santé. Cette chaîne partira de capteurs intelligents, qui produiront des signaux, suivis par des prétraitements, en particulier pour normaliser les signaux. Ensuite des caractéristiques seront extraites, puis modélisées à l'aide de modèles et de représentations. Finalement ces données seront analysées et interprétées à l'aide de connaissances expertes ou induites automatiquement.. Cette chaîne servira de base scientifique pour aborder les problématiques de santé visées par le second défi.

3.2 Verrous du défi

Les équipes du LIP6 ont choisi trois axes pour adresser les verrous de ce défi :

1. *Informatique affective*

Il s'agit de détecter les émotions ressenties par une personne, et dans certains cas de doter des agents conversationnels de la capacité à exprimer des émotions. Dans le domaine de la santé, cela peut servir à aider des personnes handicapées, dépendantes ou malades, mais aussi des étudiants sur un campus ou des travailleurs isolés en détectant, par exemple, des états préoccupants comme la tristesse, la dépression, la frustration....

Les principaux verrous sont : la modélisation des émotions ; le reconnaissance des émotions par divers canaux (voix, physiologie, texte, visage,...) ; la reproduction des émotions adaptées à l'utilisateur (surtout pour les agents conversationnels) ; les capteurs produisant des signaux adaptés à la reconnaissance des émotions.

2. *Modélisation des comportements de la personne*

Il s'agit d'analyser informatiquement le comportement d'une personne, plus spécifiquement lorsqu'il met sa santé en danger, comme par exemple, le fait de ne pas prendre les médicaments prescrits pour un diabétique, de tomber pour une personne âgée,... Pour cela, il faut identifier des informations pertinentes qui décrivent la personne impliquée dans un dispositif de e-santé pour la modéliser à plusieurs niveaux : cognitif, affectif, comportemental, intentionnel, historique des interactions, etc.

Dans le cadre de ce défi, les compétences du LIP6 permettent d'attaquer les principaux verrous suivants : introduction de nouveaux types d'information (environnement, caractéristique de la maladie, etc.) dans le modèle de la personne, et de nouveaux algorithmes de traitement adaptés (ex. modélisation par réseaux

bayésiens, estimation par filtrage optimal); construction de modèles issus de l'informatique affective pour des agents artificiels en interaction avec la personne ; suivi d'objets articulés pour analyser le comportement d'une personne dans un environnement quelconque afin d'extraire des informations comme les relations spatiales, les modèles physiques, *etc.* ; conception de capteurs intelligents qui permettent la caractérisation de la personne et de son environnement, en particulier de son état physique et émotionnel.

3. *Capteurs intelligents*

Il s'agit de concevoir des capteurs qui incluent une capacité de traitement des données. Cette capacité de traitement peut servir à améliorer les performances du capteur, par exemple comprimer les signaux pour réduire l'énergie consommée par leur transmission sans fil. Elle peut aussi servir à embarquer la chaîne d'interprétation des données dans un système ambulateur, comme une détection de chutes ou l'analyse du sommeil par exemple.

Les principaux verrous sont : les nouveaux capteurs de signaux physiologiques (en particulier pour l'informatique affective et la modélisation de la personne) et les architectures pour embarquer la chaîne d'interprétation des signaux ; l'autonomie énergétique ; l'ergonomie : capteurs sans fil,... ; la synchronisation de capteurs fournissant différentes modalités.

4 Défi : e-santé individualisée

4.1 Objectif du défi à moyen terme

L'objectif principal de ce défi est d'adapter la santé à chaque personne. Il se situe comme une application du premier défi portant sur l'interprétation des données pour la santé.

4.2 Verrous du défi

Les équipes du LIP6 ont choisi trois axes pour adresser les verrous de ce défi :

1. *Assistance*

Il s'agit de l'assistance au sens large, destinée aux apprenants, médecins, infirmiers(ères), malades, famille et aidants. Les principaux verrous sont de : développer des bases d'expériences / pratiques au-delà des simples bases de connaissances ou d'informations ; concevoir des systèmes multi-fonctionnels pour intégrer coopération, acquisition, apprentissage, explication, simulation, anticipation, formation des nouveaux référents, diagnostic collaboratif ; modéliser en temps réel des référents qui sont tous des singularités pour pallier l'absence de profils référents établis *a priori* ; réviser le type d'assistance que peut apporter un système à des référents qui sont plus experts que lui.

2. *Autonomie et maintien à domicile*

Cet axe a pour objectif la conception de technologies pour compenser une perte d'autonomie et pour accompagner la personne. Les verrous identifiés dans ce domaine sont :

l'hospitalisation à domicile et retour d'hospitalisation ; le suivi des maladies chroniques ; le suivi du vieillissement ; la compensation des déficiences ; l'accompagnement du bien-être (*coaching*).

3. *Santé personnalisée*

Il s'agit de concevoir les outils permettant de personnaliser les soins et les services apportés aux personnes.

Les principaux verrous sont : la conception de capteurs intelligents qui s'adaptent au plus près de la personne suivie ; par exemple, pour augmenter leur autonomie ; la conception d'un système d'assistance intelligent pour analyser les habitudes (et changements !) du patient ; l'aide au diagnostic (modélisation de connaissances médicales) ; prédiction de l'évolution de l'état du patient ; analyse de signaux faibles (détection d'anomalies) ; le repérage de patients répondant ou non aux modèles et la compréhension de la raison pour laquelle ils y répondent.

5 Positionnement national et international du projet : Etat de l'art / Concurrence

Le LIP6 se positionne grâce à la complémentarité des compétences et à la richesse des techniques qui ont été développées dans les différentes équipes. Plusieurs équipes possèdent d'ailleurs une expérience significative et reconnue de techniques de base et / ou de leurs applications dans ces deux défis : l'informatique

affective (MALIRE), la modélisation de la personne (ACASA, MALIRE et MOCAH). les capteurs intelligents (SYEL), l'assistance (DECISION), le maintien à domicile (PEQUAN) et la santé personnalisée (ACASA, APR et DECISION).

Les concurrents sont très nombreux : internationalement MIT, IMEC, Université en informatique de Copenhague, Université de Jyväskylä en Finlande, entreprises,...

6 Analyse SWOT

Points forts : Transversalité et synergie des compétences : Contenu (IA) + Infrastructure (HW) ; Engagement et cohésion de membres de 7 équipes déjà actives dans thème ; Compétences complémentaires présentes au LIP6 (Grandes masses données, réseau).

Points faibles : Thème appliqué, en émergence au niveau du LIP6 ; Compétences en traitement du signal et matériaux biocompatibles manquent au LIP6 ; Manque d'appui en formation initiale.

Risques liés au contexte : Concurrence : nombreux laboratoires déjà actifs. .

Possibilités liées au contexte : Equipes médicales proches au sein de l'UPMC (Pitié Salpêtrière) et de la région (AP-HP) ; Complémentarité avec ISIR et LTCI (Labex SMART) ; Importance stratégique Européenne / Nationale / Régionale / Locale de la e-santé.

7 Stratégie et mise en oeuvre

Coopération interne/ externe : D'une part, des coopérations existent entre les équipes du LIP6 actives dans le domaine de la e-santé et des partenaires médicaux. D'autre part, des coopérations internes ont émergé dans le cadre de projets internes au LIP6 : entre ACASA et SYEL sur la polysomnographie sans fil (Projet WiPo), entre MALIRE et SYEL sur les capteurs intelligents pour l'informatique affective (Projet Carpe Diem). Ceci étend des collaborations antérieures comme celle entre DECISION et PEQUAN dans le projet ANR MICO,...

Liens avec acteurs sociaux-économiques et recherche : Des liens existent avec d'autres laboratoires et établissements (ETIS, IJLRA, INRIA, IRPHE, ISIR, L2E, LIF, LSTA, LTCI, LUTIN, UDD, UTC,...), des hôpitaux (Plusieurs laboratoires -UMR 7211 CNRS-UPMC U 959 INSERM *Immunologie, Immunopathologie, Immunothérapie* - et services – de la Pitié-Salpêtrière, CHU de Rennes, Georges Pompidou, Lariboisière, Tenon,...), des organismes (Centre National de Référence pour la Santé à Domicile et l'Autonomie,...), des institutions (AP-HP, INRA, INSERM, Institut de la Longévité Charles Foix, Institut Pasteur de Tunis,...) et des entreprises actives dans le domaine de la e-santé, petites (ACACIA, CIRA, PARTELEC, PERTIMM, TRIBVN,...) ou grandes (AGFA, France-Télécom, THALES, TOPPAN Photomask,...).

Développement de compétences, qui inclut la formation : La e-santé est très peu enseignée en informatique à l'UPMC, il y a quelques UE dans le Master d'Informatique (UE Bionique de la spécialité SESI,...) et dans le Master Erasmus Mundus *Data Mining and Knowledge Management*.

Moyens existants et recherche de moyens : De nombreux projets ont eu lieu ou sont en cours au sein des équipes : locaux (Emergence UPMC ComplexIT), régionaux FEDER / FUI (Cloud-Mim, WAAVES-GP, WARM,...) ; ANR (ENDOCOM, MICO...) ; Grand Emprunt (Play and Cure,...), européens (TARDIS). Le Labex SMART inclut un axe transverse sur la e-santé que le LIP6 co-anime. Il est un lieu de synergie avec les autres thèmes scientifiques du labex SMART (modélisation de l'humain, grandes masses de données, réseau,...) et avec des laboratoires qui apportent des compétences complémentaires, comme ISIR et LTCI en traitement du signal. D'autres recherches de moyens ont été entreprises (projet DHUTMC déposé auprès de l'AP-HP en 2011, projet SESAME SMART-LAB déposé auprès de la région en 2012,...) ou vont l'être (AAP FP7 en cours).

Vague D : Campagne d'évaluation 2012-2013

2.7 Projet Transverse 2014 – 2018 - Thème SocComp

1 Présentation du thème "Social Computing"

Les données issues de réseaux, et plus particulièrement de réseaux sociaux, sont de plus en plus présentes. Leur traitement automatique représente un enjeu extrêmement fort, aussi bien du point de vue académique - la complexité de ces données nécessitent le développement de méthodes originales - que du point de vue industriel où les enjeux économiques sont très importants.

Si le domaine du traitement des réseaux sociaux est aujourd'hui très étudié dans plusieurs universités, les modèles se sont principalement intéressés au traitement de données statiques et sans contenu. Nous proposons ici de nous focaliser sur **les aspects dynamiques de ces données** : dynamiques structurelles, mais aussi dynamiques de contenu - flux d'informations, propagation d'informations, etc.

L'objectif de la prospective transverse autour du Social Computing au sein du LIP6 est de permettre aux chercheurs d'évoluer du stade d'observation et de premières pistes de compréhension des phénomènes qui se produisent dans les réseaux sociaux, au stade de la maîtrise de ces phénomènes. Dans ce cadre, plusieurs défis généraux sont à résoudre : (i) La mesure de données sociales dynamiques : comment mesurer, caractériser et quantifier la dynamique des réseaux. (ii) La modélisation : quels modèles permettent de bien simuler ce type de données. (iii) L'interaction : comment agir automatiquement avec les réseaux sociaux.

2 Projet scientifique

Le projet scientifique est structuré autour de trois problèmes clefs. Ces problèmes sont ceux de la mesure, de la modélisation et enfin de l'interaction avec les réseaux sociaux. Mis bout à bout, la résolution de ces différents points peut permettre au laboratoire d'avoir un panel complet d'outils d'analyse et de traitement des réseaux dynamiques qui n'existe pas aujourd'hui dans la communauté scientifique.

Deux principaux axes peuvent être distingués :

- La dynamique de la structure : naissance d'un réseau, organisation en communautés d'utilisateurs, modélisation de l'évolution de ces communautés et de la dynamique structurelle du réseau. Intervention sur sa structure - par l'intermédiaire d'outils de recommandation sociale par exemple.
- La dynamique des contenus : manière dont des contenus ou des propriétés se diffusent au sein d'un réseau. Comment intervenir sur la diffusion d'un message.

Le thème Social Computing réunit des chercheurs de **9 équipes et 3 départements du LIP6**. Une première force du LIP6 sur ce thème est donc le nombre et la multidisciplinarité des chercheurs impliqués. La variété des approches et des points de vue des différentes équipes sur le thème Social Computing constitue un autre atout, dans la mesure où les travaux menés par le groupe de travail ont montré la complémentarité des efforts de chacun. Ainsi, les défis décrits dans la suite de ce document tirent parti des compétences et de l'expertise de chacun en termes de collecte et d'analyse de données, de modélisation et d'optimisation. Les pistes de collaborations identifiées aujourd'hui sont les suivantes :

- Une collaboration des équipes MALIRE et BD sur l'apprentissage automatique de mesure de fiabilité des sources d'un réseau social et de popularité des données publiées.
- Une collaboration des équipes MALIRE, SMA, COMPLEX NETWORKS sur la modélisation de la diffusion d'opinions dans les réseaux sociaux dynamiques.

- Une collaboration entre APR, MALIRE et COMPLEX NETWORKS sur le développement de modèles théoriques de diffusion.
- Une collaboration entre MOCAH et COMPLEX NETWORKS concernant l'analyse des données et l'identification de communautés.
- Une collaboration possible entre les équipes DECISION, SMA, MOCAH et NPA sur l'analyse, la modélisation et la simulation des comportements dans les réseaux sociaux.
- Une collaboration avec l'équipe ACASA sur les problématiques de Social Reading et la recommandation sociale.

3 Défi Dynamique des Réseaux Sociaux : Structure, Communautés et Contenus

Ce défi fédère différents domaines de compétences présents au sein du LIP6 dans le but de lever des verrous scientifiques en lien avec la dynamique des réseaux sociaux, aussi bien du point de vue de la structure (en particulier de l'évolution de communautés d'agents), que du point de vue des contenus propagés. A partir de mesures ou de traces provenant de réseaux sociaux, il s'agit d'identifier des données qui, couplées à des règles empiriques et à des théories, permettront l'étude de comportements des individus, des communautés, et des données ou propriétés diffusées. Nous souhaitons ainsi permettre la modélisation, la simulation, la prédiction et la maîtrise des comportements de ces différentes entités. Ce défi est orienté autour de 2 axes : dynamique des structures (communautés) / dynamique des contenus.

3.1 Axe 1 : Dynamique de la structure (communautés)

Comment s'établit une communauté ? Nous cherchons, à partir de données provenant des mesures et traces collectées dynamiquement et à grande échelle dans les réseaux sociaux, à établir les liens entre les individus du réseau et à identifier des communautés. Nous souhaitons ainsi étudier sur quels critères se forment ces communautés, quelles sont leurs propriétés et leur capacité à résister à l'évolution du réseau. Une fois une communauté établie, **comment évolue-t-elle ?** Cet axe s'organise autour de l'étude de la vie de la communauté. Nous souhaitons étudier la dynamisme des communautés en cherchant à prédire les arrivées et les départs des membres de ces communautés en fonction de leurs croyances, de leurs préférences, de règles empiriques ou de théories, etc. Nous étudierons également les problématiques liées à la prise de décision collective au sein de communautés établies comme par exemple les problèmes de partage équitable, d'élicitation de préférences collectives, de coordination, d'argumentation et d'influence entre individus et communautés. Un modèle des membres d'une communauté pourrait par ailleurs être défini et enrichi au fur et à mesure de l'évolution de la communauté en traçant les activités. Des modèles comportementaux pourraient être construits pour caractériser les activités de la communauté et raisonner sur son fonctionnement. Il serait ainsi possible de détecter des leaderships, des conflits ou des comportements émergents imprévus.

3.2 Axe 2 : Dynamique des contenus

Analyse et modélisation de la diffusion dans les données sociales. Le but est, à partir de données, d'essayer de comprendre comment une information se propage dans un réseau - cela inclut notamment la diffusion d'opinions dans les réseaux sociaux. Les applications visées sont des applications autour du reporting d'activité (analyse de tendance, analyse de sentiment, analyse de l'image d'une marque dans un réseau social) ainsi que des applications multidisciplinaires telles que l'analyse des interactions sociales entre individus, ou bien l'analyse du " comportement " de maladies. De plus, nous nous intéresserons au calcul de la fiabilité et de qualité des différentes sources d'informations. **Prédiction de diffusion dans les réseaux sociaux.** Il s'agit de développer de nouvelles méthodes permettant, à l'avance, de prédire comment une ou des informations sont susceptibles de se propager. Les applications visées sont des applications de marketing viral, de détection de buzz ou de découverte de leader d'opinions. **Modèles permettant d'agir sur la diffusion.** Contrairement aux deux points précédents, nous nous intéressons ici au développement d'algorithmes qui interviennent directement sur le réseau, soit au niveau de sa structure, soit au niveau des informations transmises. Les applications visées ici sont des applications de type " contre-feu " dans le cas de la propagation de rumeur sur les réseaux sociaux. Nous pouvons aussi envisager, dans le cas de réseaux d'agents ou de réseaux informatiques, des applications de routage intelligent, la manière de router une information étant dépendante des prévisions sur le trafic à venir, ainsi

que des applications sur les protocoles P2P ou les bases de données distribuées, où chaque nJud du réseau peut décider de manière automatique quelle information propager ou non.

3.3 Verrous du défi

3.3.1 Verrous scientifiques et sociétaux Les verrous scientifiques et sociétaux sont nombreux : (i) **Identifier et Comprendre** des communautés dans un contexte dynamique, suivre leur évolution, comprendre leur fonctionnement ; (ii) **Etudier** les questions de viabilité, durabilité, résilience d'un réseau ; les processus d'argumentation au sein des réseaux et comment sont influencées les opinions des individus ; les phénomènes d'inclusion ou d'exclusion d'individus ou de groupes d'individus au sein d'une communauté (e-inclusion). (iii) **Modéliser** des processus de diffusion complexes ; l'opinion ; la fiabilité de sources de production de contenu. (iv) **Prédire** la diffusion ; le buzz

3.3.2 Verrous techniques Les verrous techniques qui se posent dans le cadre de ce défi sont étroitement liés à la taille des réseaux étudiés (i) Collecte et traitements massifs de données (ii) Passage à l'échelle, simulation à grande échelle.

3.4 Etat de l'art et Concurrence

De très nombreux travaux récents existent autour de la problématique générale du traitement de réseaux sociaux. Cependant, d'une part, ces travaux abordent peu les aspects dynamiques (structure ou/et contenu), mais de plus, ils sont très spécifiques à une tâche précise et ne sont souvent pas inséré dans un contexte de recherche plus large et plus global comme celui proposé ici. Par exemple, si l'étude des communautés dans le cas statique est aujourd'hui assez bien maîtrisée [Fortunato, 2010], l'identification et le suivi de communautés dans le cas dynamique pose de nombreux problèmes nouveaux. De même, les réseaux sociaux dynamiques sont souvent partiellement observables et les résultats des actions des individus sont donc incertains. Dans ce cadre, il est très difficile pour les agents (en terme de complexité computationnelle) de prendre des décisions de façon individuelle (décision décentralisée) qui leur permettent de maximiser un bien-être propre ou collectif [Arrow et al., 2002, Bernstein et al., 2002, Seuken et Zilberstein, 2008]. Dans le cadre de la diffusion de contenu – problématique particulièrement motivée par des applications en marketing [Bass, 1969] – les hypothèses sous-jacentes aux modèles existants – *Linear Threshold model* par exemple [Granovetter, 1978] – sont souvent peu adaptées au traitement de réseaux sociaux "du Web" et les travaux sur les données des médias sociaux sont très récents – par ex : [Leskovec et al., 2009] ou bien [Rodriguez et al., 2010].

3.5 Positionnement du projet par rapport aux défis généraux des départements

Le nombre d'équipes s'intéressant à des problématiques de diffusion est en constante augmentation. Nous citerons particulièrement l'équipe de **Jure Leskovec à Stanford** qui est l'une des plus actives dans le domaine. Cependant, aucune équipe ne regroupe simultanément les compétences permettant le traitement de tous les aspects de la diffusion (structure, opinion, contenu, stockage, etc.) contrairement au groupe multidisciplinaire que nous essayons de définir ici.

4 Analyse SWOT

Points forts : (i) Multidisciplinarité du projet et des équipes impliquées, (ii) plusieurs équipes travaillent sur ces problématiques actuellement – mais de manière séparée (iii) plusieurs collaborations actuelles avec des chercheurs nationaux et internationaux sur le sujet, ainsi qu'avec des industriels du domaine

Points faibles : (i) Problématique extrêmement active au niveau international, (ii) difficulté de la collecte et le stockage des données sociales réelles

Risques liés au contexte : (i) Gestion de l'anonymat des données, travail sur certaines données impossible, (ii) difficulté à travailler sur les données des grands industriels du domaine (Facebook, Twitter, ...) (iii) Domaine de recherche très dynamique au plan international

Possibilités liées au contexte : (i) Développer un grand groupe de travail au niveau national de recherche sur ces problématiques

5 Stratégie et mise en oeuvre

Coopération interne/ externe : Face à la très grande vitesse de la recherche internationale sur le sujet, nous envisageons la mise en place en interne de l'organisation suivante :

- Séminaire commun une fois par mois
- Réunions plus fréquentes entre les porteurs
- Réunions supplémentaires si montage de projet
- Réponse à des appels d'offre de l'ANR, de l'Europe, de la région, du pôle de compétitivité Cap Digital.

Liens avec acteurs sociaux économiques et recherche : Du point de vue des collaborations externes, nous entretenons des relations avec plusieurs industriels (Thalès, Alcatel-Lucent, BlogSpirit, ...) et académiques (Grenoble, Lille, réseau européen SINTELNET) et nous avons privilégié cet axe en faisant une proposition acceptée de projet CNRS Mastodons interdisciplinaire sur la thématique de l'analyse de données sociales. Nous sommes également leader d'un challenge international financé par le réseau d'excellence PASCAL sur la problématique de la prédiction de propagation d'informations dans les réseaux sociaux.

Développement de compétences, qui inclut la formation : Plusieurs UEs de licence et master proposées dans la prochaine maquette sont directement en rapport avec la problématique proposée :

- **Technologies du Web** (Licence 3) concerne l'étude et l'utilisation concrète de technologies récentes pour la mise en place de plateforme du Web et de réseaux sociaux.
- **Fouille de Données et Média Sociaux** (M1 Informatique) s'intéresse aux modèles de fouille dans les médias sociaux
- **Réseaux Sociaux** (Master Erasmus Mundus DNKM) : méthodes d'analyse des réseaux sociaux

Il est à noter que plusieurs UEs autour des bases de données, de l'apprentissage statistique, et des réseaux, bien que pas en rapport direct avec la problématique présentée, permettent l'enseignement d'outils indispensables pour l'analyse et le traitement de réseaux sociaux.

Moyens existants et recherche de moyens : Le LIP6 est impliqué ou porteur de plusieurs projets en cours autour des problématiques de réseaux sociaux : Projet Emergence DiRe (Diffusion) 2009-2012, Projet ANR DynGraph (Dynamique des graphes) 2010-2014, Projet FUI DIFAC (Diffusion de contenu) 2011-2014, Projet ANR MLVIS (Annotation de réseaux sociaux) 2011-2014,... Plusieurs autres projets ont été proposés récemment comme le projet FUI Amico (Recommandation sociale).

Bibliographie

[Arrow et al., 2002] Kenneth J. Arrow, Amartya K. Sen, K. Suzumura, Handbook of Social Choice and Welfare, Volume 1, Numéro 19, Handbook in Economics, North-Holland Elsevier, 2002

[Bass, 1969] F. M. Bass, A New Product Growth for Model Consumer Durables, Management Science, vol. 15, no. 5, pp. 215,227, 1969. Available : <http://dx.doi.org/10.2307/2628128>

[Bernstein et al., 2002] D. Bernstein, D., Zilberstein, S. et Immerman, N. (2002). The complexity of decentralized control of mdps. In Mathematics of Operations Research, pages 27(4) :819-840.

[Fortunato, 2010] Santo Fortunato, Community detection in graphs, Physics Reports 486, 75-174 (2010)

[Granovetter, 1978] M. Granovetter, Threshold Models of Collective Behavior, The American Journal of Sociology, vol. 83, no. 6, pp. 1420,1443, 1978. . Available : <http://dx.doi.org/10.2307/2778111>.

[Leskovec et al., 2009] J. Leskovec, L. Backstrom, and J. Kleinberg, Meme-tracking and the dynamics of the news cycle, in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ser. KDD '09. New York, NY, USA : ACM, 2009, pp.. 497,506. <http://dx.doi.org/10.1145/1557019.1557077>.

[Rodriguez et al., 2010] M. G. Rodriguez, J. Leskovec, and A. Krause, Inferring networks of diffusion and influence, in KDD '10 : Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA : ACM, 2010, pp. 1019,1028. <http://snap.stanford.edu/netinf/>

[Seuken et Zilberstein, 2008] S. Seuken, S. Zilberstein, Formal models and algorithms for decentralized decision making under uncertainty in Autonomous Agents and Multi-Agent Systems, 190-250, Volume 17-2, 2008.

Chapitre 3

Projet scientifique - Département CALSCI

Sommaire

3.1	Projet 2014 – 2018 - Département CALSCI	50
3.2	Projet 2014–2018 - Equipe PEQUAN	53
3.3	Projet 2014–2018 - Equipe POLSYS	58



Vague D : Campagne d'évaluation 2012-2013

3.1 Projet 2014 – 2018 - Département CALSCI

1 Présentation du département

Le département CALSCI regroupe les activités du LIP6 liées au calcul numérique et au calcul formel. Il se consacre au développement d'algorithmes conciliant fiabilité, robustesse et efficacité pour une variété de problèmes calculatoires. Les problèmes considérés vont de l'arithmétique de base aux systèmes algébriques non-linéaires sur des domaines continus (réels, complexes) ou discrets (corps finis). Dans le cas discret, le département a développé une expertise dans l'analyse de la sécurité des systèmes cryptographiques. Il développe plusieurs logiciels de calcul utilisés pour la résolution d'applications dans les sciences de l'ingénieur ; certaines de ces productions sont intégrées dans le système commercial MAPLE.

Le département est structuré en deux équipes. L'équipe PEQUAN se concentre sur les aspects numériques, en se focalisant particulièrement sur la fiabilité, la précision et la performance des calculs numériques. L'équipe POLSYS (commune avec INRIA Paris-Rocquencourt) se concentre sur le calcul formel qu'elle utilise pour la résolution de systèmes polynomiaux.

2 Périmètre scientifique

L'objectif central des travaux menés dans le département est de fiabiliser les outils de calcul soit en agissant sur la *methodologie* de calcul (par le calcul formel, l'usage d'arithmétiques spécifiques, etc.) soit en cherchant à *valider un calcul* obtenu (analyse *a posteriori*, méthodes stochastiques, certificats). Dans les deux cas, nous développons des algorithmes qui travaillent sur des structures de nature mathématique (nombres, matrices, polynômes, séries). Notre point de vue est que les questions de fiabilité doivent faire partie intégrante des systèmes de calcul scientifique ; notre ambition est de produire des outils fiables et performants largement visibles des communautés des sciences de l'ingénieur classiquement consommatrices de calcul scientifique, d'où le choix du nom de ce département. Nous nous sommes aussi naturellement positionnés sur des applications de notre savoir-faire dans les sciences de l'information. Notre expertise en terme de calcul dans des structures discrètes (corps finis) est mise à profit en cryptologie (attaques par canaux cachés, cryptanalyse algébrique), sujet qui est devenu un axe de recherche traité par les deux équipes avec leurs méthodologies respectives. Des recherches sont aussi menées autour du signal et de l'imagerie où interviennent des problèmes de filtrage.

3 Positionnement

Nous nous distinguons par nos approches algorithmiques que nous concrétisons par l'implantation de logiciels spécialisés à fort potentiel applicatif. Nous nous concentrons sur des problèmes d'approximation garantie, la validation numérique pour le HPC, les systèmes non-linéaires (polynomiaux) et la cryptologie. Aussi, une partie non négligeable de nos activités est tournée vers des applications.

Positionnement à l'UPMC et au LIP6. Notre potentiel applicatif s'exprime pleinement sur les aspects numériques à l'UPMC : nous collaborons avec divers laboratoires (ISIR, LOCEAN) sur des applications ainsi qu'au travers de l'Institut du calcul et de la Simulation et du LABEX CalSimLab orienté vers des applications en biologie et en chimie. Ces collaborations localisées sont illustratives de notre potentiel.

Il existe aussi des interactions potentielles fortes avec le Laboratoire Jacques Louis Lions sur des sujets plus larges ; un séminaire commun avec ce laboratoire sera mis en place pendant le prochain quinquennat afin de catalyser les collaborations possibles entre informaticiens et mathématiciens de l'UPMC et accroître notre visibilité à l'UPMC. Au LIP6, nous sommes impliqués dans les thèmes transverses *Fiabilité/Sécurité*, *Nouvelles approches en optimisation*, *Nouveaux Systèmes Répartis* et *Santé* ; ceci est plus amplement décrit

dans la section suivante. À terme nous pourrions naturellement interagir avec les départements NETSYS et SOC sur des questions liées au parallélisme et à l'arithmétique et la sécurité dans des systèmes embarqués.

Positionnement en France. Nos travaux relèvent principalement de thématiques développées au sein du GdR Informatique-Mathématiques (GT Arithmétique, Calcul Formel, Codage et Cryptographie). Soulignons aussi que des membres des deux équipes se retrouvent dans les deux derniers GT mentionnés. Via nos activités en signal, nous sommes aussi impliqués dans le GdR ISIS (Information, Signal, Images et ViSion). Ces thématiques résonnent aussi en écho à celles du GdR "Calcul" et du groupe de travail Maths-Info de la SMAI (MAIRCI). Le département collabore avec une grande majorité des équipes du thème Algorithmique, Calcul certifié et Cryptographie d'INRIA.

De nombreux laboratoires (LIENS, LIFL, LIG et LJK, LIP, LIRMM, LIX et LORIA pour ne citer qu'eux) abritent des équipes sur des thématiques similaires à celles développées dans le département. Nous collaborons avec presque toutes au travers de projets (ANR, PEPS, etc.), de thèses en co-encadrement et de publications communes. Cette activité forte dans le paysage national porte sur les aspects tant fondamentaux de nos sujets de recherche, de l'arithmétique des ordinateurs au calcul formel et à la cryptologie que sur certaines applications.

Le recrutement d'un Professeur en mutation, de deux DR INRIA et d'un CR INRIA ainsi que l'invitation de deux conférenciers du département aux journées nationales du GdR IM sont les témoins de notre attractivité et notre visibilité nationale.

Positionnement international. Nos travaux se situent dans divers groupes d'intérêt de diverses organisations internationales : l'ACM (calcul symbolique), IEEE (arithmétique), SIAM (calcul numérique et symbolique), IACR (International Association for Cryptographic Research) pour ne citer que certaines d'entre elles. Le département publie régulièrement dans les conférences et revues majeures de ces domaines, certaines d'entre elles étant organisées dans le cadre de ces organisations. Les sociétés INTEL, MAPLESOFT et MATHEMATICA sont aussi d'importants acteurs économiques consommateurs des avancées que nous produisons sur l'arithmétique, le calcul formel et le calcul numérique. Au niveau académique, le département se distingue au niveau international par la forte complémentarité des approches développées. Cela nous confère une forte visibilité et induit des collaborations avec diverses universités sur les continents asiatique, européen et nord-américains.

4 Les défis scientifiques

L'objectif ultime du département est d'impacter sur les applications tant en *sécurité de l'information et signal* que dans des *applications scientifiques pluri-disciplinaires*. Les deux leviers permettant d'atteindre cet objectif résident dans l'amélioration simultanée de l'efficacité (*calculer vite*) et de la fiabilité (*calculer mieux*) des logiciels que nous produisons. Nous donnons ci-dessous une lecture séquentielle de ces défis.

Calculer vite. Améliorer l'efficacité tant pratique que théorique de nos algorithmes et logiciels est l'un des deux verrous fondamentaux auxquels le département est systématiquement confronté. Cette quête d'efficacité recouvre plusieurs objectifs : gagner un facteur constant sur des opérations de base (un facteur 2 sur un opérateur arithmétique a un impact immédiat sur tous les codes l'utilisant), ou gagner plusieurs ordres de grandeur sur des spécifications sophistiquées pour rendre accessible des calculs qui ne l'étaient pas.

Pour atteindre ces deux objectifs, les verrous à lever sont doubles. D'une part, un problème est d'exploiter au mieux les nouvelles architectures (multi-cœur, GPU, *manycore*, etc.) qui offrent un ordre de grandeur supplémentaire en puissance de calcul, mais qui sont aussi parfois porteuses de nouvelles contraintes (parallélisme massif et régulier, gestion des données, contraintes des systèmes embarqués). Un second verrou réside dans les complexités théoriques parfois exponentielles des problèmes que nous cherchons à résoudre (certains d'entre eux étant situés aux frontières de la décidabilité).

Pour attaquer le premier verrou, nous remodelons en amont les algorithmes afin d'obtenir des implantations efficaces ; nous exploitons pour cela notre savoir-faire en arithmétique et en algorithmique algébrique et numérique. Pour attaquer le second verrou, nous cherchons évidemment à diminuer au maximum les complexités, y compris au niveau des constantes (parfois situées en exposant) lorsque celles-ci peuvent être identifiées ; nous ne nous intéressons donc pas uniquement aux complexités asymptotiques. Un autre angle d'attaque qui a émergé durant le dernier quinquennat et qui devrait s'amplifier est l'exploitation des structures apparaissant fréquemment dans les problèmes que nous devons résoudre (systèmes linéaires ou systèmes polynomiaux structurés) pour obtenir des algorithmes dédiés significativement plus rapides.

Le thème transverse *Nouveaux Systèmes Répartis* offre ici l'opportunité de collaborer sur des aspects système et architecture en s'appuyant sur nos algorithmes, logiciels et problèmes applicatifs pour concevoir, intégrer ou valider des environnements d'exécution ou des spécificités matérielles.

Calculer mieux. Comme indiqué plus haut, un enjeu pour le département est de concilier fiabilité et efficacité.

Il faut souligner ici les besoins croissants de certification :

- du point de vue numérique, la conception de bibliothèques générant automatiquement des approximations fiables pour divers opérateurs ou fonctions spéciales reste un problème largement ouvert ;
- du point de vue du calcul exact, l’usage d’algorithmes probabilistes et la technicité des implantations affaiblissent potentiellement le niveau de confiance dans les résultats produits (notamment, lorsqu’un ensemble vide est retourné pour un problème de décision d’existence de solutions)

Du point de vue des spécifications, le département concentrera ses études sur des problèmes non-linéaires comme des problèmes de décision (existence de solutions pour les systèmes polynomiaux) et d’approximation (évaluation polynomiale, génération automatique de fonctions spéciales, etc.). Les verrous sont triples : calculer avec une plus grande précision, avoir une information sur la précision d’un calcul, calculer des certificats.

Les leviers que nous développerons pour attaquer ces verrous sont l’usage de calcul formel (y compris pour des problèmes d’approximation numérique), de l’analyse de précisions et de constructions algébriques pour calculer des certificats de non-existence de solutions.

Ces travaux s’inscrivent dans la thématique transverse “Nouvelles approches en optimisation” via la conception d’algorithmes conciliant certification et performances pour l’optimisation dans des contextes non-linéaires. Les applications scientifiques de ces algorithmes sont importantes, d’où notre investissement dans ce thème.

Sécurité de l’information. En cryptologie, un objectif est de développer les interactions entre PEQUAN et POLSYS dans le domaine spécifique de l’analyse de fuites d’informations. L’ambition est de concevoir des opérateurs arithmétiques (domaine d’expertises de PEQUAN) qui soient à la fois efficaces et sécurisés c’est à dire qui ne laissent pas fuir d’informations (par exemple une dysymétrie dans un calcul engendre une variation mesurable de consommation électriques). Pour garantir la sécurité du système, ces primitives doivent résister aux meilleures attaques connues. Parmi celles-ci, les attaques algébriques (domaine d’expertise de POLSYS) permettent de retrouver à partir d’un petit nombre de mesures, parfois indirectes (par exemple le poids de Hamming d’une valeur d’état interne), la clé secrète du cryptosystème. Ces perspectives de recherche s’inscrivent dans le thème transverse “Fiabilité/Sécurité” : un des défis est d’aborder les questions d’attaques et de fuites d’informations sous l’angle à la fois matériel (ALSOC/CIAN), logiciel (MOVE) et théorique.

Applications scientifiques pluri-disciplinaires. Une partie des activités de recherche du département CALSCI est menée dans le cadre de collaborations pluridisciplinaires avec des partenaires académiques ou industriels. Nos domaines d’application sont très variés : aéronautique, astro-physique, biologie, chimie quantique, mécanique des fluides, physique atomique, sciences du climat, robotique, théorie du contrôle et théorie du signal...

Ces applications permettent de valider les avancées algorithmiques et d’améliorer les logiciels développés dans le département et La complémentarité des travaux du département s’illustre parfaitement dans le cadre de ces travaux : l’équipe PEQUAN intervenant surtout sur les applications relevant de la simulation alors que l’équipe POLSYS intervient sur les applications relevant de la conception.

5 Evaluation du département : analyse SWOT

Points forts. La complémentarité et l’homogénéité thématique du département sont des leviers potentiels pour impacter fortement sur certains sujets. Le département dispose aussi d’une expertise de premier plan dans la résolution des systèmes polynomiaux et la validation numérique. Les résultats obtenus sur le calcul de bases de Gröbner sont remarquables et apportent une très forte visibilité internationale au département.

Points faibles. Certaines thématiques sont très instables du fait du manque de taille critique et qu’elles soient portées par des Maîtres de Conférences ayant soutenu leurs HDR très récemment (ou sur le point de la soutenir). Ce manque de taille critique est un frein pour aller au-delà de notre complémentarité et obtenir des productions communes à fort impact.

Risques liés au contexte. L’investissement des rangs A du département dans les tâches collectives (resp. Licence et Master d’Informatique, resp. PolyTech’UPMC) impacte lourdement sur l’encadrement des activités de recherche ; cette situation est critique notamment pour l’équipe PEQUAN.

Possibilités liées au contexte. Notre investissement dans l’Institut du Calcul et de la Simulation, du LABEX CalSimLab et la mise en place d’un séminaire commun avec le LJLL sont autant de leviers dont nous disposons pour nous permettre d’impacter dans le monde du calcul scientifique. De même, dans le contexte INRIA qui est une des tutelles de l’équipe POLSYS nous travaillons régulièrement avec presque toutes les équipes du thème *Algorithmique, Calcul certifié et Cryptographie*. Enfin, du côté des enseignements, le département est impliqué dans le montage d’une spécialité *Fiabilité, Sécurité et Performances du Numérique*.



agence d'évaluation de la recherche
et de l'enseignement supérieur

Unités de recherche

Vague D : Campagne d'évaluation 2012-2013

3.2 Projet 2014–2018 - Equipe PEQUAN

1 Membres permanent de l'équipe

Nom	Prénom	Status	Etablissement
Bajard	Jean-Claude	PU1	UPMC
Bézeziat	Dominique	MC	UPMC
Chesneaux	Jean-Marie	PU0	UPMC
Dubuisson	Séverine	MC	UPMC
Fortin	Pierre	MC	UPMC
Graillat	Stef	MC	UPMC
Hilaire	Thibault	MC	UPMC
Jezequel	Fabienne	MC	UNIV PARIS 2
Lamotte	Jean-Luc	PU2	UPMC
Lauter	Christoph	MC	UPMC
Ménissier-Morain	Valérie	MC	UPMC

2 Evolution de l'équipe

La création au sein de l'UPMC de l'institut du calcul et de la simulation en janvier 2012 accompagné d'un LabEx CalSimLab et d'un équipex Equip@Meso crée un appel à collaboration avec des utilisateurs du calcul intensif comme la chimie, la biologie et la physique, l'océanographie et bien d'autres disciplines. Nous avons obtenu un an de financement de la part du LabEx sur un projet intitulé "Astrophysical N-body simulations on multi-core and many-core architectures". De fait, l'équipe veut renforcer son axe calcul parallèle haute performance avec la demande d'un poste de professeur qui viendra soutenir notre action en direction de l'institut du calcul et de la simulation.

3 Caractérisation de la recherche

Le positionnement de PEQUAN se situe en recherche fondamentale et recherche appliquée.

D'un côté le travail de recherche de PEQUAN sur la représentation des données numériques et des algorithmes associés, ainsi que sur la validation et les paradigmes du calcul distribué entrent dans le cadre d'une recherche que l'on peut qualifier de fondamentale. De l'autre, PEQUAN met en application ses études théoriques pour offrir des outils et des développements donnant lieu soit à une mise à disposition "Open Source", soit à des études intéressant l'industrie. La force de PEQUAN est de pouvoir offrir des applications assises sur des bases théoriques solides.

4 Projet scientifique

Evolutions par rapport à la période précédente /Prise de risque L'équipe PEQUAN est en plein essor, son activité s'est fortement accrue avec l'arrivée de financements de projets au cœur de ses thématiques,

qui lui permettent de mener à bien son activité de recherche en terme de publication et d'encadrement d'étudiants en thèse ou en master.

Positionnement par rapport aux thèmes transverses et de département Les quatre thèmes présents au sein de PEQUAN collaborent étroitement. Tous utilisent des architectures parallèles GPU ou multi-core qui sont le cœur de métier du thème "calcul haute performance". La fiabilité et la validation numérique du thème 1, fait la force et l'originalité des travaux menés sur le thème 4.2. De même, nos connaissances en représentation des données numériques nous permettent de les mettre en application en fonction des particularités des filtres. L'imbrication des thèmes est un point fort de l'équipe, et donne lieu à des publications communes.

Enfin, l'équipe participe à la vie du département sur deux axes transverses qui sont la cryptographie et le calcul symbolique numérique.

4.1 Thème "Précision et fiabilité des calculs"

Description

Génération automatique de codes numériques L'équipe PEQUAN compte exploiter et approfondir son expertise en termes de gestion de précision et de fiabilité des calculs, acquise dans l'ANR TaMaDi et la bibliothèque Metalibm, pour automatiser la génération de codes d'évaluation de fonctions, de filtres et de régulateurs. La génération automatique de code ouvre aussi la voie à des optimisations de codes numériques avec une approche originale : il s'agit à la compilation de remplacer l'analyse sémantique classique du code par un échantillonnage numérique.

Certification symbolique-numérique et calcul multi-précision validé Tout au long des étapes de génération automatique, des certificats doivent être produits pour justifier de la correction numérique du code produit. L'équipe PEQUAN compte, pour traiter ce problème, poursuivre son investissement sur la calcul symbolique-numérique en collaboration avec l'équipe PolSys.

Petite Grande Précision Les générateurs automatiques auront des besoins localisés en précision de calcul plus grande que ce qui est offert par le matériel (double précision), sans que le calcul tout entier bascule dans une arithmétique multi-précision comme MPFR qui serait bien plus coûteuse. Il y a aussi une forte demande au sein de la communauté pour un support logiciel en Petite Grande Précision.

Originalité L'équipe PEQUAN réunit à présent des compétences variées dans le domaine de la génération automatique de codes numériques allant des fonctions élémentaires aux filtres pour le traitement du signal, en se focalisant dans les deux cas sur la précision des calculs. La réunion de ces deux composantes au sein de l'équipe PEQUAN est tout à fait unique en France et va créer des synergies originales et permettre des avancées significatives.

Positionnement dans le contexte local/ national/ international La génération de code engendre une multitude de problèmes d'optimisation qui devrait ouvrir à des collaborations au sein du LIP6. La génération de codes logiciels pour des fonctions mathématiques nécessite un savoir-faire d'optimisation pour l'évaluation de polynômes, pour laquelle l'équipe DALI de l'université de Perpignan est experte et avec qui nous envisageons une coopération. Quant à l'outil FloPoCo de l'ÉNS Lyon, il réalise une génération automatique d'implantations de fonctions numériques mais pour une cible matérielle les FPGA, l'approche est donc très différente. La compagnie RedHat Inc. nous a fait part de son vif intérêt et de sa volonté de coopérer sur nos travaux de générations de codes pour les fonctions élémentaires qui est, à notre connaissance, un thème non développé à l'étranger.

Objectifs de résultats à 5 ans Pour la génération automatique d'implantations de fonctions mathématiques que l'on trouve dans une libm, l'équipe envisage de livrer une bibliothèque paramétrée ainsi qu'un environnement de test. Ce logiciel s'appuiera uniquement sur des outils Open Source afin d'intégrer la bibliothèque libm Open Source de glibc.

Nous souhaitons étendre cette génération automatique à l'implantation de filtres et de régulateurs, avec le souci de la précision des calculs.

L'équipe PEQUAN souhaite décliner un ensemble d'outils logiciels pour la Petite grande précision. avec avant tout un souci d'efficacité. Nous envisageons de développer une bibliothèque de *quad precision open source* satisfaisant la norme IEEE754-2008, avec un souci premier de complétude et de fiabilité.

4.2 Thème "Calcul à Haute Performance"

Description Sur le thème HPC (High Performance Computing), l'équipe PEQUAN compte poursuivre son travail sur la conception d'algorithmes numériques optimisés en temps d'exécution pour diverses architectures parallèles et sur la validation numérique de logiciels scientifiques. Elle compte investir le champs des applications réelles de très grande taille qui tendent à devenir cruciales dans plus en plus de domaines.

Originalité Une des spécificités de l'équipe PEQUAN est sa double expertise en conception d'algorithmes numériques performants et en validation numérique de logiciels.

Positionnement dans le contexte local/ national/ international Dans le cadre de la validation numérique, l'arithmétique d'intervalles est peu adaptée au cas des gros codes industriels (surestimation de l'erreur générée avec les algorithmes numériques classiques), et le calcul d'estimateurs de conditionnement présente aussi certaines limites (restriction à certaines classes de problèmes, surestimation de l'erreur).

L'arithmétique stochastique, développée dans l'équipe PEQUAN, implique plusieurs exécutions du code effectuées de manière synchrone en utilisant un mode d'arrondi aléatoire et nécessite peu de modifications du code cible : elle permet ainsi la validation numérique de logiciels scientifiques de grande taille.

Objectifs de résultats à 5 ans Grâce aux contacts industriels que nous avons développés, nous envisageons de contrôler la qualité numérique d'applications HPC de grande taille. Nous souhaitons mêler différentes méthodes de validation numérique (arithmétique stochastique, estimateurs de conditionnement) dans une bibliothèque parallèle d'analyse d'erreur. Comme les moyens de calcul évoluent, il sera indispensable d'adapter les méthodes de validation numérique aux nouvelles architectures, en veillant à ce que le coût de ces méthodes reste acceptable dans un contexte industriel.

Nous souhaitons aussi prolonger et développer nos collaborations actuelles (académiques et industrielles) sur l'algorithmique numérique pour le HPC, en visant notamment l'adéquation matériel-logiciel et portabilité des performance, qui sont et resteront des enjeux cruciaux pour de nombreuses applications.

4.3 Thème "Filtres et Images"

Description Nous souhaitons étudier l'impact des calculs en précision finie pour les algorithmes de traitement d'images 2D et 2D+t déjà développés dans le thème.

Originalité Le portage de nos algorithmes de traitements de signaux nD a déjà largement été étudié par la communauté scientifique pour ce qui concerne les GPU et autres architectures HPC. En revanche, les aspects précision et fiabilité des calculs n'ont pas encore vraiment été étudiés pour le cas des architectures embarquées (par exemple micro-contrôleur ou FPGA).

Positionnement dans le contexte local/ national/ international À notre connaissance, il n'y a pas d'études dans l'état de l'art sur le problème de la précision des calculs dans le contexte que nous visons. Ces travaux, à la croisée de la vision et de la fiabilité des calculs auront ainsi un positionnement très original dans le paysage actuel.

Objectifs de résultats à 5 ans La vidéo surveillance est un domaine en pleine extension, avec non seulement des applications de sécurité, mais aussi médicales (surveillance sanitaire de personnes âgées par exemple). La masse de données à traiter est immense, et une partie du traitement (la détection d'événements d'intérêt précis – la chute d'une personne par exemple) pourrait être déportée sur l'appareil d'acquisition, réduisant ainsi le coût de stockage et d'analyse de ces données sur le serveur central. Par contre cela demande de porter sur ces périphériques une partie des calculs numériques et de s'assurer de leur fiabilité. Ce thème s'inscrit également dans une des perspectives transverses, portant sur la E-santé, du laboratoire.

4.4 Thème "Opérateurs pour la cryptographie"

Description La mise au point d'une arithmétique résistante aux fautes et aux attaques par analyse des fuites est un véritable challenge dans le monde de la sécurité.

Originalité Actuellement, la plupart des contre-mesures faites pour lutter contre les attaques sont intimement liées au support sur lequel est implanté le crypto-système. L'originalité de notre approche est de jouer directement sur le codage des données et les algorithmes associés.

Positionnement dans le contexte local/ national/ international L'équipe PEQUAN est reconnue mondialement pour son approche utilisant des systèmes résiduels "Residue Number Systems" (RNS). Elle est la seule équipe française experte dans le domaine.

Objectifs de résultats à 5 ans Dans le cadre de la bourse DGA de Julien Eynard, nous nous sommes engagés à extraire des propriétés du RNS la possibilité de résister aux fautes. Nous désirons faire une analyse des fuites via des outils de mesure de la qualité de l'aléa en terme de sécurité (tests du NIST par exemple). Enfin, dernier challenge, les similitudes entre les RNS et l'interpolation polynomiale sont nombreuses (Lagrange, Newton,...), nous aimerions trouver des bases permettant d'approcher les transformées rapides de Fourier.

5 Analyse SWOT

Points forts : PEQUAN regroupe un savoir faire assez large autour de l'arithmétique des ordinateurs qui va de la validation numérique sur les flottants, à l'arithmétique en multiprécision, des opérateurs pour l'embarqué en virgule fixe pour le traitement du signal ou exact pour la cryptographie, au calcul numérique sur grandes architectures parallèles ainsi que sur GPU. Ce partage de connaissances lors des groupes de travail hebdomadaires, donne à notre équipe une grande expertise du calcul sur ordinateurs.

Points faibles : Les trois professeurs de l'équipe ont de lourdes charges dans l'administration de l'enseignement : un directeur d'école d'ingénieur, un directeur de département de licence et un directeur de département de master.

Possibilités liées au contexte : Nous constatons un intérêt croissant de l'industrie pour les thèmes développés au sein de PEQUAN. Au niveau du calcul parallèle des entreprises comme TOTAL, EDF, AMD nous ont contactés pour mener certaines études. De même, en cryptographie, des sociétés comme ST-Microelectronics ou Cassidian (EADS) souhaitent collaborer avec nous.

Risques liés au contexte : Ces domaines demandent une forte réactivité liée aux besoins de l'environnement mais aussi en lien avec la concurrence mondiale sur le sujet. L'offre de nouvelles architectures, ainsi que l'apparition de nouveaux protocoles ou de nouvelles méthodes numériques créent de nouveaux challenges auxquels nous nous devons de répondre si nous voulons rester dans la course.

6 Mise en œuvre

Valorisation et recherche de moyens : Partenariat, montage de projets

Soumission de projets à venir :

- Validation numérique d'applications HPC en mécanique des fluides (avec LRI-ONERA)
- Génération automatique de codes parallèles et fiables pour l'assimilation de données (avec LOCEAN-LRI-Télécom Sud Paris)
- Soumission d'un projet ANR INS "Arithmétiques Robustes et Randomisées pour des Architectures Sécurisées" avec STMicroelectronics, Cassidian, univ. Grenoble, univ. Rennes.

Diffusion des résultats (stratégie de publication) : Conférence de référence vs. workshops

Stratégie locale / nationale/ Internationale : UPMC-Region / LABEX - Animation communauté / Actions internationales : projets, réseaux

- *Participation à l'Institut du Calcul et de la Simulation de l'UPMC et au labex CalSimLab* portant sur la modélisation et la simulation numérique des grands défis de la chimie et de la biologie (projet interdisciplinaire entre mathématiciens, informaticiens, physiciens, chimistes et biologistes). En particulier pour PEQUAN : projet de déploiement sur architectures hétérogènes (massivement) parallèles de méthodes hiérarchiques pour le problème à N-corps en astrophysique et en dynamique moléculaire.
- *Participation au dépôt de projet européen EMC2 (Expanding Mathematics and Computational Chemistry, <http://www.ics.upmc.fr/modules/resources/download/ics/ERC/emc2.html>).*

Implication future dans l'enseignement : Lien avec la nouvelle maquette du L et M, Echanges internationaux

Le département CALSCI (PEQUAN et PolSys) sont porteurs, dans le cadre du master d'informatique, d'une nouvelle spécialité "Sécurité, Fiabilité et Performance du Numérique" (SFPN) qui reprend en partie des UE existantes en les réactualisant vers les thèmes Calcul Parallèle, Calcul Formel, Cryptographie, Fiabilité numérique. L'équipe PEQUAN interviendra dans 5 UE : "Modélisations numériques et symboliques", "Programmation parallèle", "Représentation des Nombres et Algorithmique", "Calcul Haute Performance" et "Arithmétique Flottante et Analyse d'Erreurs". D'autre part, l'équipe offre en Licence 2e année un cours d'introduction aux méthodes numériques.



Vague D : Campagne d'évaluation 2012-2013

3.3 Projet 2014–2018 - Equipe POLSYS

1 Membres permanent de l'équipe

Nom	Prénom	Status	Établissement
Berthomieu	Jérémy	MC	UPMC
Faugère	Jean-Charles	DR	INRIA
Perret	Ludovic	MC	UPMC
Renault	Guénaël	MC	UPMC
Safey El Din	Mohab	PR	UPMC
Tsigaridas	Elias	CR1	INRIA
Wang	Dongming	DR	CNRS

2 Évolution de l'équipe

L'équipe-projet POLSYS¹ (Polynomial Systems) commune entre INRIA et l'UPMC fait suite à l'équipe-projet SALSA. Une évolution notable est que nous plaçons maintenant au coeur de notre recherche le calcul de bases de Gröbner ainsi que l'utilisation des bases de Gröbner comme brique de base des algorithmes de résolution des équations polynomiales dans les corps finis ou sur les réels.

E. Tsigaridas, recruté CR1 (INRIA) au 1-er septembre 2012, renforcera l'activité de l'équipe concernant la résolution des systèmes polynomiaux sur les réels. J. Berthomieu, recruté MC (UPMC) au 1-er septembre 2012, travaillera sur des algorithmes utilisés en amont car permettant de simplifier les systèmes polynomiaux en vue du calcul de base de Gröbner.

3 Caractérisation de la recherche

Nous étudions le problème fondamental de la recherche de solutions d'équations algébriques. Les problèmes non linéaires apparaissent fréquemment dans de nombreux domaines des sciences de l'ingénieur ainsi qu'en Théorie de l'information (Cryptographie, Théorie du Signal, ...). Dans l'équipe, nous proposons des algorithmes et des analyses de complexité et nous utilisons les applications pour valider l'efficacité des implantations de ces algorithmes. Certaines applications nous contraignent à développer, dans un premier temps, des méthodes adhoc que nous généralisons ensuite. Une part importante de notre activité consiste à écrire des logiciels efficaces que nous utilisons dans les applications.

4 Projet scientifique - Défis

À long terme, l'objectif principal du projet est de repousser significativement les limites actuelles des algorithmes de résolution des systèmes polynomiaux et leur utilisation dans les applications. Le verrou principal provient de la complexité *intrinsèquement* exponentielle de ces problèmes. Nous proposons d'atteindre cet objectif en abordant la question sous plusieurs angles : amélioration des algorithmes fondamentaux (même

1. Une description complète du projet est aussi disponible : <http://www-polsys.lip6.fr/~jcf/Papers/PolSys.pdf>

lorsqu'ils ont une complexité exponentielle), développement d'implantations haute-performance intégrant le parallélisme, et recherche de nouvelles classes de problèmes algébriques *structurés* qui peuvent être résolus en temps polynomial. L'objectif est aussi de résoudre des problèmes difficiles issus de domaines variés comme la Cryptologie, la théorie des nombres (corps finis), géométrie algébrique réelle (élimination des quantificateurs), problèmes d'optimisation globale et géométrie algorithmique.

Le calcul efficace des bases de Gröbner prend une place importante dans notre nouveau projet scientifique. C'est en effet notre moteur de calcul principal, qui permet de réécrire les systèmes donnés en entrée en des systèmes équivalents dotés de propriétés exploitées par des routines de plus haut niveau, soit pour la résolution dans les corps finis (applications en Cryptologie), soit pour la résolution sur les nombres réels (applications dans les sciences de l'ingénieur). L'élaboration du projet scientifique de POLSYS se fonde sur l'identification de défis et de nouvelles tendances du domaine :

- *Algorithmes efficaces pour les problèmes NP-difficiles.* Sur les corps finis la recherche de solutions des systèmes polynômes est un problème NP-difficile et on ne peut espérer obtenir des algorithmes sous-exponentiels. Ceci ne signifie pas qu'on ne peut pas améliorer la complexité : ainsi dans le cas d'un algorithme simplement exponentiel de complexité $O(2^{cn})$, la possibilité de réduire significativement la constante c peut avoir un impact important sur la taille des problèmes qui peuvent être accessibles en pratique.
- *Systèmes Structurés ou comment contourner la complexité exponentielle.* Si la complexité de la résolution est au moins simplement exponentielle dans le cas générique, la plupart des problèmes provenant des applications possèdent une structure supplémentaire. Le défi est alors d'identifier des classes de problèmes pour lesquels la complexité est polynomiale en le nombre de solutions et donc de proposer des algorithmes dont la complexité dépend de la taille de la sortie.
- *Algorithmes pour la géométrie algébrique réelle efficaces en théorie en pratique.* De nombreuses applications des sciences de l'ingénieur nécessitent de trouver ou caractériser les solutions *réelles* de systèmes polynomiaux. En pratique, les algorithmes qui sont implantés dans les systèmes de Calcul Formel sont récursifs (projections successives) ce qui induit une complexité doublement exponentielle en le nombre de variables. D'un point de vue théorique, il existe des algorithmes de complexité simplement exponentielle qui répondent à la plupart de ces questions. Cependant, ces algorithmes ne sont pas praticables et ne sont souvent pas implantés. Le défi est donc de réduire le fossé entre théorie et pratique. On peut ainsi espérer des avancées significatives et d'atteindre des applications qui sont pour l'instant hors de portée.
- *Cryptanalyse Algébrique – comprendre la complexité.* La *Cryptanalyse Algébrique* est maintenant une méthode bien établie pour analyser la sécurité d'un cryptosystème. Elle peut se décomposer en deux étapes : une phase de modélisation du problème par des équations algébriques puis une phase de résolution (souvent un calcul de base de Gröbner). Une tendance nouvelle du domaine est d'associer aux expérimentations pratiques une analyse théorique de complexité. Le défi est donc d'utiliser les propriétés particulières de ces systèmes polynômes pour expliquer la complexité de l'attaque. Cette question est liée à la complexité de résolution des systèmes structurés.

La proposition de projet est architecturée autour de quatre objectifs. Même si ce nombre est important au regard de la taille du groupe il faut remarquer que ces objectifs ne sont pas indépendants. Par exemple, tout progrès réalisé dans le cadre de l'objectif 4.1 ou 4.3 aura immédiatement un impact significatif sur les autres objectifs.

Positionnement au LIP6. Localement, ce projet entre en résonance avec plusieurs thématiques transverses du Laboratoire. Mentionnons tout d'abord la thématique "fiabilité et sécurité informatique" à laquelle nous apportons une expertise en cryptologie (notamment en cryptanalyse algébrique). Nous nous positionnons également sur la thématique liée à l'optimisation via notre activité sur les systèmes structurés et la résolution des systèmes sur les réels.

Positionnement national et international. Au niveau national, l'équipe collabore régulièrement avec les équipes du thème "Algorithmique et calcul certifié" d'INRIA, divers groupes de cryptologie et les équipes universitaires de calcul formel. Son positionnement, bien centré sur la résolution des systèmes polynomiaux par des méthodes exactes, alliant développements théoriques, logiciels et résolution d'applications est original et nous confère une bonne visibilité. Au niveau international, nous sommes la seule équipe à regrouper en son sein des experts du domaine sur tous les angles mentionnés ci-dessous ; notre visibilité et notre implication dans la communauté internationale est renforcée par nos nombreuses collaborations, ainsi que notre animation

d'équipes associées internationales.

4.1 Objectif 1 : Algorithmes fondamentaux et Systèmes Structurés.

L'objectif est d'une part de proposer des algorithmes efficaces (souvent de complexité exponentielle) pour les opérations de base (calcul de base de Gröbner, isolation des racines réelles) et d'autre part d'identifier de large classe de problèmes structurés que l'on peut résoudre en temps polynômial.

Autres méthodes pour résoudre les systèmes polynômiaux : Ensembles Caractéristiques généralisés. Même si le calcul des bases de Gröbner restera notre activité de recherche centrale nous prévoyons aussi de porter notre effort dans les méthodes alternatives. En particulier l'équipe possède une expertise forte dans les méthodes fondées sur les ensembles triangulaires.

4.2 Objectif 2 : Solutions réelles des systèmes polynômiaux et Applications.

Pour les systèmes algébriques en dimension positive (c'est à dire lorsqu'il y a un nombre infini de solutions complexes) les questions les plus basiques peuvent être très difficiles (par exemple le problème de tester l'existence de solutions réelles). Nous prévoyons de proposer des algorithmes et des implantations afin de répondre aux questions les plus importantes : calcul de points tests représentant un ensemble infini de solutions réelles, décider si deux tels points sont dans la même composante connexe et, à plus long terme, l'élimination des quantificateurs sur les réels (c'est à dire qu'à partir d'une formule quantifiée impliquant des équations et des inéquations polynômiales on peut calculer une formule équivalente sans quantificateur). Un objectif est de concilier meilleures performances pratiques et classes de complexité asymptotiquement optimales.

4.3 Objectif 3 : Algèbre linéaire dédiée pour le calcul des bases de Gröbner. Implantations HPC.

L'efficacité des algorithmes de calcul des bases de Gröbner comme F_4 et F_5 reposent fortement sur l'efficacité de l'algèbre linéaire. En effet, ces algorithmes procèdent par réduction à des problèmes d'algèbre linéaire (élimination de Gauß sur des matrices de grandes tailles). L'objectif est de tirer parti de la structure des matrices (quasi-triangulaire par blocs ou caractère creux) pour proposer des algorithmes d'algèbre linéaire dédiés. Du point de vue logiciel l'objectif est le développement d'une bibliothèque d'algèbre linéaire haute performance (multicore et parallèle) ; le mode de développement privilégié est un développement collaboratif sous license LGPL.

4.4 Objectif 4 : Résolution dans les corps finis. Applications en cryptologie et théorie des nombres.

Nous proposons une utilisation systématique des résultats sur les systèmes structurés en cryptanalyse algébrique. L'objectif est d'améliorer l'efficacité pratique et de prédire la complexité théorique de ces attaques. Une piste pour améliorer l'efficacité pratique est – par exemple – de tenir compte du fait qu'on cherche les solutions dans un corps fini. Nous proposons d'étudier la complexité de la méthode hybride (mixant calcul de base de Gröbner et recherche exhaustive). Nous pensons aussi étendre les attaques algébriques dans de nouveaux domaines de la cryptographie comme la théorie algébrique des nombres (typiquement, la cryptographie fondée sur les courbes).

Au delà des cryptanalyses (attaques), nous pensons aussi utiliser nos outils algébriques pour estimer finement les paramètres de sécurité nécessaires pour certains cryptosystèmes. Plus récemment, nous avons examiné la possibilité d'utiliser la difficulté du problème de résolution des systèmes polynômiaux (et des variantes bruitées de ces problèmes classiques) pour concevoir des cryptosystèmes.

5 Analyse SWOT

Points forts : POLSYS est une équipe leader mondiale dans le domaine de la résolution des systèmes polynômiaux avec un spectre allant de la théorie (résultats de complexité) à la pratique (impact dans un domaine comme la cryptographie) avec des développements logiciels importants (Maple) permettant une activité de transfert technologique. Nous avons un taux de publication dans les conférences et revues de références important par rapport à la taille de l'équipe.

Points faibles : La taille de l'équipe peut être vue comme un point faible. Il est prévisible que plusieurs MdC ayant une forte expertise en cryptographie soutiennent des HDRs pendant la prochaine période d'évaluation. Cette partie de l'activité de recherche de POLSYS pourrait être fragilisée.

Risques liés au contexte : Le non-remplacement de MdC travaillant en cryptologie est un risque pour cette activité importante de l'équipe.

Possibilités liées au contexte : Au sein de l'INRIA nous collaborons avec de nombreuses équipes (aussi bien sur les aspects Calcul Formel que sur la cryptographie). Sur le plan industriel nous pouvons collaborer avec des acteurs majeurs du domaine de la sécurité (Thalès, Oberthur, Gemalto, DGA, ...).

6 Mise en œuvre

Valorisation et recherche de moyens : Partenariat, montage de projets

Diffusion des résultats (stratégie de publication) : Nous avons mis en place une stratégie de publication : pour les activités liées à notre cœur de métier, le Calcul Formel nous privilégions les publications dans les conférences et revues principales du domaine (ISSAC, Journal Of Symbolic Computation, ...). Pour les applications nous faisons le choix de publier dans les conférences ou revue de référence de ces domaines (Crypto, Eurocrypt, Asiacrypt, SOCG, DCG, ...).

Stratégie locale / nationale/ Internationale : UPMC-Region / LABEX - Animation communauté / Actions internationales : projets, réseaux

L'équipe POLSYS organise plusieurs conférences internationales (SCC, Macis,...) et nous organisons l'édition de plusieurs numéros spéciaux qui aident à animer scientifiquement la communauté. Nous prévoyons de développer les relations avec la côte est des USA avec le soutien de l'équipe associée avec l'Université de Caroline du Nord.

Implication future dans l'enseignement : Lien avec la nouvelle maquette du L et M, Echanges internationaux

L'équipe est fortement impliquée dans les formations de Licence et Master de l'UPMC. Elle y développe naturellement des enseignements à l'interface entre informatique et enseignement comme la cryptologie. Au niveau Licence, un de ses membres assure la co-responsabilité du parcours "Informatique, Mathématiques et Applications"; dans le cadre de ce parcours, un séjour d'études à l'étranger est organisé. L'équipe assure aussi la responsabilité de plusieurs Unités d'Enseignement; notamment autour de thématiques liées à la cryptologie et l'algorithmique. Au niveau Master, l'équipe assure la responsabilité de plusieurs Unités d'Enseignement autour de thématiques liées à la cryptologie, la sécurité et le calcul formel. Elle intervient aussi au MPRI (Master Parisien de la Recherche en Informatique). Un membre de l'équipe est responsable d'une spécialité en cours de création au sein du Master d'Informatique de Paris 6 (Sécurité, Fiabilité et Performance Numérique).

Chapitre 4

Projet scientifique - Département DESIR

Sommaire

4.1	Projet 2014 – 2018 - Département DESIR	64
4.2	Projet 2014–2018 - Equipe RO	68
4.3	Projet 2014–2018 - Equipe DECISION	72
4.4	Projet 2014–2018 - Equipe SMA	77
4.5	Projet 2014–2018 - Equipe MOCAH	82



Vague D : Campagne d'évaluation 2012-2013

4.1 Projet 2014 – 2018 - Département DESIR

1 Présentation du département

Le département DESIR (“DECision, Systèmes Intelligents et Recherche opérationnelle”) du LIP6 rassemble 35 permanents (11 professeurs ou directeurs de recherche, 23 maîtres de conférences ou chargés de recherche) et 51 doctorants qui travaillent au développement de modèles et d’algorithmes pour la représentation de problèmes complexes et de leur résolution par la machine, éventuellement en interaction avec un ou plusieurs agents humains. Nous travaillons également à la conception de systèmes intelligents (agents adaptatifs, systèmes multi-agents, systèmes d’aide à la décision, d’aide à la formation) et à leur mise en oeuvre au sein d’applications industrielles.

Les membres du département se répartissent en 4 équipes qui sont spécialisées dans le domaine de la *Recherche Opérationnelle* (responsable Ch. Dürr, 11 permanents), de la *Décision* (responsable P. Perny, 12 permanents), des *Systèmes Multi-Agents* (responsable A. El Fallah Seghrouchni, 6 permanents) et des *Modèles et Outils en ingénierie des Connaissances pour l’Apprentissage Humain* (responsable J. M. Labat, 6 permanents). Les interactions entre ces domaines sont multiples et sont à l’origine de collaborations étroites entre nos équipes, qu’il s’agisse de recherche fondamentale ou d’applications industrielles. Nous collaborons également au niveau de l’enseignement en préparant conjointement la spécialité ANDROIDE (AgeNts Distribués, Robotique, Recherche Opérationnelle, Interaction, DECision) dans le cadre de la nouvelle maquette du master informatique de l’UPMC.

2 Périmètre scientifique

Les activités du département relèvent de l’intelligence artificielle, de l’aide à la décision et de la recherche opérationnelle. Elles portent sur la modélisation et l’optimisation des systèmes, l’aide à la décision et la décision automatique, les agents adaptatifs et les systèmes multi-agents, les systèmes interactifs et de formation.

Nous cherchons à couvrir un large spectre d’activités incluant des aspects fondamentaux (modélisation mathématique, analyse axiomatique, étude de complexité), des aspects algorithmiques et programmation pour la résolution de problèmes et le développement de systèmes intelligents (agents adaptatifs, système multiagents, systèmes d’aide à la décision, d’aide à la formation) et des aspects applicatifs liés à leur mise en oeuvre au sein d’applications industrielles. Les compétences principales des équipes portent sur les thèmes suivants :

- RO : recherche opérationnelle, algorithmique, optimisation combinatoire, ordonnancement, satisfiabilité, routage dans les réseaux,
- DECISION : décision multicritère, décision collective, décision dans l’incertain et le risque, optimisation en nombres entiers, optimisation dans les réseaux, systèmes d’aide à la décision, contexte,
- SMA : apprentissage multi-agents, coordination, décision distribuée et planification multi-agent, dialogue et interaction, simulation,
- MOCAH : environnements interactifs d’apprentissage humain, modélisation de l’apprenant, diagnostic cognitif, outils auteurs, serious games.

Outre ces sujets traditionnellement présents, nous cherchons à développer des axes originaux réunissant plusieurs équipes sur des problématiques récentes, par exemple au carrefour de l’optimisation et de la théorie de la décision (théorie de la décision algorithmique), des systèmes multiagents, de l’algorithmique et des jeux

(théorie des jeux algorithmique), de l’algorithmique et du choix social (choix social computationnel), de la représentation des connaissances et des jeux sérieux.

3 Positionnement

Une des spécificités du département DESIR réside dans sa double culture Recherche Opérationnelle et Intelligence Artificielle et son positionnement intermédiaire entre recherche fondamentale et recherche appliquée. Les travaux menés dans le département portent sur des problèmes formels qui traduisent des situations complexes du monde réel d’aujourd’hui et de demain et visent à apporter des solutions sans recourir à des simplifications drastiques. Ainsi, la complexité croissante des problèmes d’organisation, de contrôle, d’évaluation et de diagnostic, d’optimisation et de décision qui se posent dans le domaine de la gestion de l’énergie, de l’optimisation de l’espace et du temps, des transports et des procédés, de l’information et la communication, de la santé, de l’élaboration de politiques publiques et de la gouvernance, nous conduisent à étudier de nouveaux problèmes et à travailler au développement de nouveaux outils. En particulier nous cherchons à concevoir :

- des méthodes d’optimisation ou d’inférence capables de traiter des instances de très grande taille et de les résoudre de manière exacte ou approchée en fournissant une garantie sur la qualité de la solution ou leur robustesse,
- des systèmes de décision capables de déterminer les meilleures options dans des environnements de plus en plus complexes (incertitude, multi-critère, multi-agent, contexte distribué, contexte stratégique),
- des systèmes intelligents impliquant des agents capables de s’adapter aux évolutions de leur environnement, d’apprendre collectivement et de coopérer pour résoudre des problèmes,
- des environnements interactifs d’apprentissage humain reposant sur une modélisation de l’apprenant et permettant un diagnostic cognitif.

Ces travaux rejoignent certaines des préoccupations énoncées dans les actions transverses envisagées par l’ensemble du LIP6 pour les prochaines années et devraient donc permettre d’augmenter les interactions avec d’autres équipes du laboratoire. Les équipes du département sont en effet concernées par les actions transverses portant sur les *nouvelles approches de l’optimisation*, sur *l’intelligence ambiante*, sur le *social computing* et sur la *e-santé*. Par ailleurs le versant IA du département et en particulier les travaux autour des agents, de la décision et la modélisation des préférences, de l’analyse du contexte, du e-learning et des systèmes interactifs s’insèrent naturellement dans le LABEX SMART, notamment sur l’axe “Digital and Human Reasoning”. Au niveau national et international, nos équipes sont largement impliquées dans la vie de l’AFIA et de la ROADEF, dans la vie et l’animation des GDR RO, du groupe “SMA et Simulation” du GdR I3 et dans le GDR International AlgoDec.

4 Les défis scientifiques

Sans recourir à un catalogue exhaustif, nous listons ci-dessous quelques défis spécifiques, représentatifs de nos objectifs et préoccupations pour les prochaines années et des thèmes susceptibles de bénéficier d’interactions entre plusieurs équipes du département.

Nouveaux défis pour l’ordonnancement. De nouveaux contextes d’exécution engendrent des problèmes d’ordonnancement, qui ont soit de nouvelles contraintes à respecter soit de nouvelles fonctions objectif à optimiser. Concrètement les calculs sur les cartes graphiques GPU posent de tout nouveaux problèmes d’ordonnancement. Un autre aspect est celui de la gestion de la température et de la vitesse de calcul qui n’a été que récemment pris en compte. En particulier la minimisation de l’énergie consommée est un nouveau critère d’optimisation en ordonnancement et en lot-sizing, qui génère des problèmes de nature nouvelle.

Programmation mathématique. La programmation mathématique a été développée avec le domaine de l’optimisation au milieu du 20^{ème} siècle. L’approche primale duale a été à la base de l’optimisation combinatoire et des algorithmes d’approximation. Cet outil peut aussi être développé pour l’algorithmique en ligne. Beaucoup de techniques issues de l’analyse polyédrale comme la génération de coupes n’ont pas encore été appliquées aux algorithmes d’approximation, et ceci constitue une piste prometteuse que nous allons engager. Par ailleurs, concernant la résolution exacte de problèmes d’optimisation en nombres entiers difficiles (réseaux

télécom, logistique, transport) nous chercherons d'une part à développer de nouveaux types de relaxations par reformulation-linéarisation et d'autre part de mener l'étude de polyèdres combinatoires particuliers.

Optimisation combinatoire multicritère. Un des défis actuels consiste à développer des méthodes générales d'identification de l'ensemble de Pareto dans des problèmes combinatoires admettant plus de deux critères. Une autre voie, complémentaire, consistera pour nous à chercher des algorithmes efficaces pour construire des approximations de l'ensemble de Pareto avec garantie de performance. On cherchera enfin à approcher des compromis spécifiques par optimisation d'une fonction d'agrégation non-linéaire.

Décision collective, coordination et mechanism design. Le développement de modèles, algorithmes et protocoles pour la prise de décision multiagents comporte de nombreux défis, actuellement à l'étude dans la communauté : comment s'accommoder de capacités de communications limitées, d'informations incomplètes, imprécises, voire volontairement manipulées par les agents, tout en garantissant des solutions acceptables individuellement et satisfaisantes collectivement ? Nous mobilisons ici des outils de la décision dans l'incertain (comme les DecMDPs), du choix social computationnel, ou encore de la théorie des jeux pour aborder ces questions. Par ailleurs, la coordination multi-agents invoque souvent des décisions collectives et requiert des protocoles multi-agents capables de prendre en compte des critères qualitatifs (ex. intérêts des agents, bien-être collectif, équité ou atteinte de consensus) et des critères quantifiant l'efficacité de la coordination (ex. fiabilité des communications, coûts, asynchronisme). Nous nous intéresserons à l'intégration simultanée de ces deux types de critères dans la conception de mécanismes de coordination.

Modèles graphiques. Qu'il s'agisse de réseaux Bayésiens, de réseaux de préférences, de réseaux de contraintes, ou de jeux graphiques, les modèles graphiques s'imposent actuellement comme un formalisme puissant pour représenter de manière compacte des croyances, des coûts ou des préférences sur des espace combinatoires. Nous chercherons à développer cet axe présent dans le département et à produire des algorithmes originaux exploitant les structures graphiques pour résoudre des problèmes d'inférence ou d'optimisation de manière plus efficace.

Jeu vidéo et jeu sérieux. Si de nombreux jeux vidéo sont multi-joueurs et même souvent massivement multijoueurs, il n'en va pas de même pour les jeux sérieux qui sont la plupart du temps des jeux mono-joueurs car il est difficile de contrôler l'apprentissage de chacun dans un jeu multi-joueurs. Un objectif est de définir de nouvelles stratégies d'apprentissage basées sur la coopération et/ou la compétition dans les jeux et pour cela d'utiliser les approches, les langages et les protocoles de communication conçus au sein de l'équipe SMA.

Modélisation de l'utilisateur. Nous prévoyons d'aborder la question de l'adaptation du système informatique à l'humain, une question difficile mentionnée dans beaucoup de domaines, par exemple dans le champ de l'Intelligence Ambiante ou des Objets Communicants. Cette adaptation repose en premier lieu sur un diagnostic (cognitif, voire affectif) des humains, sur des modèles fins de décision dans l'incertain et souvent multi-critères et sur la capacité des agents logiciels à modifier dynamiquement leur comportement pour aider ou anticiper sur les attentes des utilisateurs.

5 Evaluation du département : analyse SWOT

Points forts. La complémentarité des compétences présentes au sein du département favorise la formulation de nouveaux problèmes et leur résolution et participe au développement de nouveaux thèmes transverses au niveau national et international. Une des forces du département est la variété des techniques d'analyse et de résolution utilisées pour résoudre les problèmes difficiles que nous abordons. Nos activités couvrent un spectre large qui va du fondamental à l'expérimental. Enfin le département présente une bonne attractivité pour les chercheurs (1 DR et 2 CR CNRS recrutés ces dernières années, plusieurs visiteurs de renommée internationale).

Points faibles. Les recrutements d'enseignants-chercheurs ont permis de combler les départs et renouveler les compétences mais pas d'augmenter l'effectif malgré le développement de nos activités. De plus, nous manquons d'un ingénieur de recherche pour développer et maintenir des plateformes logicielles. Enfin, certaines interactions restent à développer au sein du département et avec d'autres départements du laboratoire.

Risques liés au contexte. Concurrence au niveau national et international dans plusieurs communautés (IA, RO, informatique théorique, économie, théorie des jeux), existence de laboratoire entièrement focalisés sur nos thématiques.

Possibilités liées au contexte. Bon réseau d'équipes en interaction, que ce soit sur les aspects décision/RO (LAMSADE, CEDRIC, LIX, CERMSEM, LINA, LIPN, LRI) ou sur les aspects IA, (LAMSADE, LTCI, ISIR, IRIT, CRIL, LIPN, CRIP5, GREYC, MAGMA, LIFL, LAMIH). Développement du Labex SMART à l'UPMC et de diverses actions transverses au LIP6 impliquant notre équipe. Les problématiques abordées concernent les industriels (optimisation des systèmes, aide à la décision, intelligence ambiante, formation).



Vague D : Campagne d'évaluation 2012-2013

4.2 Projet 2014–2018 - Equipe RO

1 Membres permanent de l'équipe

Nom	Prénom	Statut	Etablissement
Angelopoulos	Spyros	CR	CNRS
Bampis	Evipridis	PR	UPMC
Delbot	François	MC	Univ. Paris-10
Dubois	Olivier	CR	CNRS
Dürr	Christoph	DR	CNRS
Fouilhoux	Pierre	MC	UPMC
Hanen	Claire	PR	Univ. Paris-10
Hyon	Emmanuel	MC	Univ. Paris-10
Kedad-Sidhoum	Safia	MC	UPMC
Pascual	Fanny	MC	UPMC

2 Evolution de l'équipe

Après le départ de Francis Sourd, le départ en retraite de Philippe Chrétienne et l'arrivée de Evripidis Bampis, Spyros Angelopoulos, Christoph Dürr et François Delbot, l'équipe s'est ouverte à de nouveaux sujets en algorithmique, tout en développant des travaux ambitieux en recherche opérationnelle. Nous espérons pouvoir compléter ce groupe par un poste de professeur d'ici 2018 aussi pour combler les départs en retraite dans le département, en particulier en programmation mathématique.

3 Caractérisation de la recherche

Notre équipe a la particularité de travailler à la fois sur des projets appliqués et des projets théoriques. Cette proximité génère des discussions fructueuses, et permet à chacun d'envisager de nouvelles problématiques ou d'aborder un problème avec de nouvelles techniques.

4 Projet scientifique

Nous avons organisé notre projet scientifique autour de trois thèmes, qui sont l'analyse fine d'algorithmes et complexité, l'ordonnancement dans de nouveaux contextes et l'optimisation en environnement coopératif et concurrentiel. Ainsi une partie de notre projet s'inscrit dans l'optimisation, un thème transversal au laboratoire. De plus, la théorie des jeux algorithmique, et la programmation mathématique sont des domaines dans lesquels de fortes interactions existent avec les autres membres du département.

4.1 Analyse fine d'algorithmes et Complexité

4.1.1 Phénomène de seuil Nous projetons de développer l'étude du phénomène de seuil pour le problème de couverture d'un ensemble par des sous-ensembles (Exact Cover). Ce problème très simple dans sa définition,

permet d'étudier le phénomène en fonction d'un nombre réduit de paramètres ce qui devrait permettre de mieux préciser les conditions qui le produisent. Nous entreprendrons également l'étude du phénomène de seuil sur les systèmes linéaires aléatoires sur un corps fini. C'est une extension à des domaines non booléens, de dimension quelconque non abordés jusqu'à présent dans les phénomènes de seuil. Cette étude devrait en particulier apporter un nouvel éclairage sur les limitations des techniques utilisées et les moyens d'y remédier. On peut par ailleurs en attendre des retombées pratiques concernant le hachage et les codes correcteurs d'erreurs.

4.1.2 Analyse en moyenne des algorithmes approchés L'analyse en moyenne est une mesure complémentaire à l'analyse au pire de cas, permettant d'obtenir des résultats plus fins que ceux rendus possibles par la seule évaluation en pire cas. Que l'on se focalise sur la qualité des solutions retournées ou sur le temps d'exécution, il sera nécessaire de s'intéresser au comportement des algorithmes mais également à la structure des instances manipulées. Ce type d'étude devrait permettre une meilleure compréhension de la dynamique des algorithmes d'approximation et ainsi favoriser la conception d'algorithmes plus performants.

4.1.3 Nouvelles techniques d'analyse des algorithmes en ligne Dans cette thématique nous cherchons à développer de nouvelles mesures, modèles et techniques pour l'optimisation en ligne. Actuellement l'algorithmique en ligne est centrée sur l'analyse dans le pire des cas, qui ne reflète malheureusement pas toujours les performances observées en pratique. En particulier nous avons besoin d'outils qui permettent d'expliquer pourquoi certains algorithmes se comportent mieux en pratique que d'autres. Nous nous tournons alors vers de nouvelles techniques d'analyse qui ont été tout récemment proposées (comme l'analyse bijective), de nouveaux modèles plus fidèles au monde réel (comme des modèles pour la pagination avec principe de localité) et de nouvelles approches pour l'analyse (comme l'analyse stochastique remplaçant l'analyse au pire cas). Cette recherche s'insère dans un programme ANR "New Techniques in Online Computing" (2011-2015).

4.1.4 Programmation mathématique Nous aimerions comprendre comment l'analyse polyédrale peut aider l'analyse des algorithmes. Par exemple récemment une approche primale-duale a fourni un environnement unificateur pour de nombreux algorithmes en ligne, comme l'atteste le livre ["The Design of Competitive Online Algorithms via a Primal-Dual Approach" par Niv Buchbinder and Seffi Naor, 2009]. Une autre approche liant programmation mathématique et algorithme d'approximation serait d'étudier comment dériver des algorithmes d'approximation à partir d'algorithmes de coupes.

4.2 Ordonnancement dans de nouveaux contextes

4.2.1 Robustesse et incertitudes Considérer que les tâches ne sont plus déterministes mais soumises à des aléas est un des axes en plein essor de l'ordonnancement. On pourra chercher soit un ordonnancement robuste dont la sensibilité à l'aléa sera la plus faible possible soit dans un cadre entièrement stochastique chercher des stratégies de contrôle optimales. Les axes qu'il nous semble intéressant d'explorer sont d'une part de travailler sur le sens de la robustesse et de sa définition mathématique associée ; d'autre part, d'évaluer les performances et la complexité de ces algorithmes dans des cas classiques de l'ordonnancement. L'autre approche en contrôle optimal concerne la recherche de propriétés structurelles du contrôle dont la connaissance évite une énumération de l'ensemble des états et diminue la complexité du calcul.

4.2.2 Ordonnancement cyclique Les problèmes d'ordonnancement cyclique dont les tâches doivent être répétées indéfiniment, font l'objet de méthodes de résolution approchées de qualité déjà bien établies. Mais les questions d'avenir, comme celles liées à la minimisation de la consommation d'énergie, ou la prise en compte de contraintes temporelles comme les latences négatives traversent également ce domaine. L'équipe RO a les compétences pour les étudier et proposer des algorithmes efficaces, en poursuivant la collaboration initiée avec l'université de Prague. De plus, certaines questions plus fondamentales laissées ouvertes par nos recherches, par exemple la qualité d'approximation fournie par les ordonnancements périodiques dans les graphes d'événements généralisés seront explorées.

4.2.3 Ordonnancement sur les nouvelles plate-formes de calcul Dans le cadre du soutien Google-CNRS et en collaboration avec Denis Trystram du LIG, nous développons actuellement un nouvel axe de recherche

en ordonnancement pour les nouvelles plates-formes de calcul avec critères non-standards. Cette nouvelle thématique bénéficie du soutien du GDR-RO en 2012. L'objectif est de proposer des approches génériques pour des architectures modernes — notamment celles disposant de processeurs CPU et d'accélérateurs GPU — capables de prendre en compte les nouvelles caractéristiques de ces systèmes puis d'étudier des algorithmes d'ordonnancement pour ces modèles dans les différents cas de figures.

4.2.4 Ordonnancement pour minimiser l'énergie consommée pendant le calcul Notre objectif premier est de considérer des modèles plus évolués dans lesquels on prendrait en compte l'hétérogénéité de processeurs. Quand on parle d'hétérogénéité dans ce contexte on souhaite exprimer le fait que les processeurs ont des fonctions de puissances différentes, ce qui implique que l'exécution des parties d'une même tâche avec la même vitesse sur deux processeurs différents induit une consommation d'énergie différente, ce qui rend le problème beaucoup plus complexe et traduit les architectures modernes mariant CPU et GPU p.ex.

Nous projetons d'étudier des problèmes en ligne dans ce cadre en cherchant à utiliser l'approche primale duale par exemple. L'aspect décentralisé de certains systèmes nous amènent à étudier ces problèmes en introduisant des aspects de la théorie des jeux.

4.2.5 Recherche de cible dans des rayons Dans cette classe de problèmes, un explorateur mobile doit localiser une cible qui se trouve dans une position qui lui est inconnue, dans un environnement qui lui peut être connu ou inconnu à son tour. Le but est de concevoir des algorithmes d'exploration qui trouvent la cible le plus rapidement possible. Ceci modélise, par exemple, l'optimisation dans les forages pour la recherche de pétrole ou l'exécution de tâches parallèles dont les durées sont inconnues. Ce domaine a déjà été très étudié, cependant de nombreuses questions importantes n'ont pas encore été abordées. Celles-ci incluent par exemple la localisation conjointe de plusieurs cibles, que nous avons commencé à étudier dans le contexte de cibles localisées sur plusieurs rayons. Une autre direction est de définir de nouvelles mesures du coût de l'algorithme qui soient plus pertinentes. Finalement on cherchera à étudier le modèle plus général, quand l'algorithme connaît une distribution probabiliste des positions possibles des cibles. Ces problèmes ne sont pas seulement intéressants d'un point de vue théorique, mais ont de belles applications en intelligence artificielle et en recherche opérationnelle.

4.3 Optimisation en environnement coopératif ou concurrentiel

4.3.1 Optimisation conjointe et optimisation multi-critères L'étude des problèmes conjoints apparaît aujourd'hui comme naturelle dans le cadre de la résolution de problèmes réels de recherche opérationnelle. Ils consistent à considérer en même temps les aspects parfois contradictoires des solutions aux problèmes industriels qui combinent des prises de décisions diverses : affectations, localisation, routage,... Une approche prometteuse est l'étude des polyèdres résultant de l'union des polyèdres associés à chacun des aspects du problème, au travers par exemple de formulations étendues. L'optimisation de problèmes conjoints est particulièrement importante dans les problématiques d'ordonnancement et de planification. Il serait intéressant d'utiliser ces approches pour les problèmes comme le RCPSP (*Resource-Constrained Project Scheduling Problem*) ou le cadre large de l'optimisation robuste.

4.3.2 Théorie des jeux algorithmique Nous souhaitons continuer à étudier la conception d'algorithmes ayant pour entrée des données appartenant à des utilisateurs aux intérêts divergents, ces algorithmes devant retourner une solution de bonne qualité (vis à vis d'une fonction de coût social) et acceptable par tous les utilisateurs. D'une part, nous souhaitons analyser plus finement les algorithmes que l'on concevra en comparant la solution qu'ils retournent non pas à la meilleure solution pour le coût social, mais à la meilleure solution qui respecte la contrainte d'acceptabilité par les utilisateurs. D'autre part, nous souhaitons rapprocher ces travaux de la théorie des jeux coopératifs. En particulier nous sommes intéressés par des solutions dans lesquelles il n'est pas (ou peu) intéressant pour les utilisateurs de créer des coalitions qui leur permettent de refuser la solution retournée par l'algorithme. Et finalement nous chercherons à utiliser des algorithmes de résolution issus de la programmation linéaire pour améliorer les techniques de calcul effectif existantes et également d'utiliser des outils issus du multicritère pour élargir les critères d'équité.

4.3.3 Planification de production avec contraintes environnementales Nous souhaitons poursuivre les travaux initiés autour des chaînes logistiques vertes pour lesquels nous bénéficions d'un projet PEPS CNRS (2011-2012) en collaboration avec le G-Scop et le CMP-EMSE. Nous avons conçu plusieurs modèles innovants en intégrant les aspects environnementaux sous forme de contraintes, plutôt que de composante de coût dans la fonction objectif comme dans la majorité des travaux déjà existants. De nombreuses pistes sont à l'étude, nous souhaitons en particulier étudier les problèmes qui traitent de quantités approvisionnées entières et d'autre part ceux qui intègrent des contraintes de capacité.

5 Analyse SWOT

Points forts : Une force de l'équipe RO est la variété des techniques d'analyse et de résolution utilisées pour résoudre des problèmes difficiles. Une partie des problèmes de recherche puisent leurs sources dans des applications industrielles. Les thématiques abordées restent ainsi pour la plupart concrètes et bénéficient pour leur résolution d'outils théoriques pointus. Ceci permet à l'équipe de rester attractive auprès du monde économique et de proposer des travaux reconnus dans le domaine académique.

Points faibles : Les outils et thématiques portés par les nouveaux permanents de l'équipe ouvrent de nouvelles perspectives de collaboration qui devraient se concrétiser à travers des projets communs, même si à ce jour il y a encore peu de publications communes. Il est à noter qu'un séminaire actif offre actuellement un cadre d'échanges enrichissant entre les membres de l'équipe.

Risques liés au contexte : Comme nous projetons de travailler sur des sujets porteurs actuellement, cette situation nous plonge dans une communauté internationale active et riche, mais nous place aussi dans un contexte de concurrence accrue imposant une exigence plus forte sur les résultats obtenus.

Possibilités liées au contexte : Depuis une dizaine d'années la recherche opérationnelle est devenue un domaine académique reconnu, et l'équipe bénéficie du réseau qui s'est créé autour du GdR RO et de la conférence ROADEF. Aussi depuis quelques années il y a une nouvelle vague de chercheurs ambitieux en algorithmique, et notre équipe bénéficie de la proximité de laboratoires parisiens comme le LIAFA, le LIENS, le Cédric, et le LAMSADE.

6 Mise en œuvre

Valorisation et recherche de moyens : Nous cherchons des financements à la fois publics — comme p.ex. dans le 7th Framework Programme de la communauté européenne — et privés — comme p.ex. EDF.

Diffusion des résultats (stratégie de publication) : Nous cherchons à communiquer dans les colloques importants de nos communautés (Workshop on Lot-Sizing, Workshop on Models and Algorithms for Planning and Scheduling Problems, etc.) et aspirons à l'excellence pour le choix des journaux et conférences (SIAM Journal on Discrete Mathematics, STOC, ESA, etc.)

Stratégie locale / nationale/ Internationale : Nous cherchons à maintenir notre présence dans le réseau national et international en recherche opérationnelle (Journées franciliennes de recherche opérationnelle, European Operational Research Society, Gotha, etc.) et à aider à développer la communauté algorithmique en France (groupe de travail complexité et algorithmes, etc.).

Implication future dans l'enseignement : Nous allons poursuivre l'enseignement de la recherche opérationnelle et de l'optimisation combinatoire à l'UPMC et dans des Masters, comme le MPRI, et cherchons à être présent dans le MPRO (Cnam-ParisTech, Ecole Polytechnique, Telecom, ENSTA, ensiie).



Vague D : Campagne d'évaluation 2012-2013

4.3 Projet 2014–2018 - Equipe DECISION

1 Membres permanents de l'équipe

Nom	Prénom	Statut	Etablissement
Isabelle	Alvarez	ICPEF	IRSTEA
Patrick	Brézillon	PU2	UPMC
Christophe	Gonzales	PU2	UPMC
Thibault	Lust	MC	UPMC
Michel	Minoux	PU0	UPMC
Viet Hung	Nguyen	MC	UPMC
Hacène	Ouzia	MC	UPMC
Patrice	Perny	PU0	UPMC
Olivier	Spanjaard	MC	UPMC
Paolo	Viappiani	CR1	CNRS
Paul	Weng	MC	UPMC
Pierre-Henri	Wuillemin	MC	UPMC

2 Evolution de l'équipe

A la rentrée 2012, un CR1 CNRS a intégré l'équipe décision en la personne de Paolo Viappiani. Ses travaux sur l'élicitation de préférences et les systèmes de recommandation s'intègrent naturellement dans nos activités et viennent les compléter. Nous venons également de recruter Thibault Lust sur un poste de maître de conférences pour la rentrée 2012. Formé à l'optimisation combinatoire multi-objectifs dans le service de Mathématique et Recherche Opérationnelle de la faculté polytechnique de Mons, ses recherches s'intégreront naturellement dans le thème 1 (décision sur domaine combinatoire).

Dans un futur proche, nous souhaiterions recruter un maître de conférences travaillant sur les modèles graphiques (réseaux Bayésiens, réseaux de préférences) pour augmenter notre potentiel de développement de nouveaux algorithmes pour l'inférence et l'optimisation et notre capacité d'intervention sur des applications. Par ailleurs, le recrutement d'un spécialiste de théorie des jeux algorithmique et/ou du choix social computationnel nous paraît souhaitable pour développer les aspects stratégiques et algorithmiques de la décision collective (manipulation, mechanism design) qui prennent de l'importance actuellement en IA et décision. Nous chercherons également à recruter un professeur pour prendre progressivement la suite des activités de Michel Minoux, actuellement professeur émérite, dans l'animation du thème 3 (modélisation et optimisation des systèmes) et pour continuer de développer les applications de la programmation mathématique et de l'optimisation combinatoire à la théorie de la décision algorithmique (thème 1).

3 Caractérisation de la recherche

Les travaux de l'équipe couvrent un spectre assez large qui va de préoccupations fondamentales concernant la justification théorique des modèles ou des algorithmes à des activités plus appliquées concernant la résolution

pratique de problèmes de décision complexes ou le développement de systèmes de décision. L'effort porte principalement sur la modélisation de problèmes de décision ou d'optimisation en environnement complexe (définition des solutions potentielles, modélisation des préférences, de l'incertitude), l'étude de la complexité de problèmes, le développement de méthodes de résolution exactes ou approchées avec garantie de performance pour le calcul des décisions optimales, l'aide au diagnostic et l'explication des décisions.

4 Projet scientifique

Le projet de l'équipe est d'oeuvrer au développement de la théorie de la décision algorithmique et pour cela de continuer l'effort entrepris concernant le développement des modèles et des algorithmes pour la prise de décision en environnement complexe. Nous présentons les principales directions dans lesquelles nos trois thèmes d'activité vont évoluer.

4.1 Thème 1 : décision sur domaine combinatoire, optimisation multi-objectifs

Description Ces dernières années, l'équipe a contribué, avec quelques autres en France et à l'étranger, à l'émergence de la *théorie de la décision algorithmique*, un courant récent en théorie de la décision qui consiste non plus seulement à travailler sur les modèles mathématiques et leur justification axiomatique, mais aussi sur les aspects représentation machine des problèmes de décision et sur le calcul des solutions préférées. Ce changement de perspective motivé par la taille des problèmes à résoudre et la nature combinatoire des espaces de solutions à explorer attire un nombre croissant d'informaticiens qui s'attaquent à des problèmes algorithmiques nouveaux qui résultent de l'utilisation de modèles décisionnels sophistiqués sur des espaces de solutions trop grands pour être défini de manière explicite. L'équipe prévoit de continuer à oeuvrer au développement de travaux au carrefour de la théorie de la décision et de l'optimisation combinatoire. Outre la poursuite des travaux en cours en optimisation combinatoire multicritère ou multi-agents, en optimisation robuste et en décision dynamique dans l'incertain nous prévoyons de développer des aspects relevant du choix social computationnel (vote sur domaine combinatoire), de la théorie des jeux algorithmique (manipulation de procédure de décision et conception de mécanismes à vérité garantie), thèmes qui connaissent un succès croissant actuellement au delà de la sphère décisionnelle. Quelques travaux dans cette direction sont déjà présents dans l'équipe et annoncent cette évolution. Certains font d'ailleurs l'objet d'interactions avec les équipes RO et SMA du département DESIR et devraient se développer au sein du département.

Originalité L'originalité de ce thème est de revisiter l'optimisation combinatoire sous l'angle de la théorie de la décision pour aborder des problèmes où la recherche d'une solution optimale est guidée par des structures de préférences complexes (multi-critères, non-complètes, non-linéaires), nécessitant une algorithmique spécifique exploitant les propriétés des modèles décisionnels impliqués. L'équipe décision qui réunit des compétences à la fois sur les fondements mathématiques et axiomatiques des modèles décisionnels (quantitatifs et qualitatifs), sur les outils de la recherche opérationnelle (algorithmique des graphes et programmation mathématique) et de l'Intelligence Artificielle (représentation des connaissances, des préférences et des croyances) pour la modélisation et la résolution de problèmes est bien configurée pour aborder ces problématiques originales portant sur les aspects computationnels de la théorie de la décision.

Positionnement dans le contexte local/ national/ international Au niveau local et en particulier au sein du département DESIR, ces problématiques rejoignent en partie celles des équipes RO (optimisation combinatoire, théorie des jeux algorithmique) et celles de l'équipe SMA (décision multi-agents, décision distribuée) et sont un sujet naturel de collaboration. Sur le plan national et international l'équipe décision est l'une des équipes moteur en théorie de la décision algorithmique, en interaction avec les équipes actives sur ce sujet au LAMSADE, à l'IRIT, au CRIL, au LINA et au GREYC, qu'il s'agisse de projets ANR ou de collaborations plus ponctuelles, mais aussi avec les équipes participant au GDR International AlgoDec récemment créé sur ces thèmes.

Objectifs de résultats à 5 ans Formulation de nouveaux problèmes de décision sur domaines combinatoires (multicritères, multi-agents, sous-incertitude), étude de leur complexité, production d'algorithmes exacts et approchés pour le calcul des solutions préférées, implémentation et tests numériques.

4.2 Thème 2 : Modèles graphiques pour le raisonnement dans l'incertain et la décision

Description L'équipe a entamé des recherches sur l'exploitation de modèles graphiques pour les systèmes industriels complexes et de grandes tailles. Dans ce cadre, nous poursuivrons nos efforts sur l'inférence probabiliste structurée, permettant ainsi de passer à l'échelle. De plus, nous étudierons de nouvelles requêtes probabilistes comme la détection d'événements rares ou inhabituels.

Nous répondrons également à de nouvelles problématiques industrielles en proposant des algorithmes d'apprentissage de réseaux probabilistes "généraux" à partir de grandes masses de données non structurées et potentiellement hétérogènes. Ceux-ci seront capables de passer à l'échelle sans trop restreindre le pouvoir descriptif des réseaux obtenus (e.g., réseaux bayésiens dynamiques non stationnaires).

Enfin, en ce qui concerne les graphes contextuels, leur identification à une base d'expériences nous amène maintenant à étendre les fonctionnalités du logiciel CxG (par exemple, en permettant de lancer des actions extérieures au logiciel). Cela nous permettra de poser les fondations des systèmes d'assistance intelligents basés sur le contexte qui sont au cœur de nos préoccupations pour les années à venir.

Originalité Pour les réseaux bayésiens, l'inférence structurée et l'apprentissage sans contraintes à partir de grandes masses de données sont encore un verrou technologique. Pour les graphes contextuels, notre approche vise à remplacer le concept de base de connaissances par celui de base d'expériences.

Positionnement dans le contexte local/ national/ international L'équipe possède actuellement les algorithmes d'inférence structurée parmi les plus performants au niveau international et veut conserver ce leadership. En proposant un cadre conceptuel robuste, nos travaux sur le contexte se départissent fortement des travaux à l'international.

Objectifs de résultats à 5 ans Les objectifs sont, d'une part, de produire une chaîne complète d'algorithmes pour résoudre des problèmes décisionnels de grandes tailles et, d'autre part, d'établir les fondations des systèmes d'assistance intelligents basés sur le contexte.

4.3 Thème 3 : Modélisation et Optimisation des Systèmes

Description Il s'agit d'étudier et de mettre en oeuvre de nouveaux types de relaxations utiles en vue de la résolution exacte et approchée de problèmes d'optimisation en nombres entiers difficiles rencontrés dans de nombreuses applications industrielles telles que réseaux de télécommunications, réseaux logistiques, problèmes de transports, etc. Plus précisément, dans le prolongement des travaux actuels de l'équipe, les principaux axes sur lesquels nous allons porter notre effort sont les suivants :

- a) concernant les relaxations par reformulation-linéarisation, on étudiera d'une part des représentations compactes, d'autre part la possibilité de renforcements utilisant des modèles de programmation semi-définie ; parmi les applications visées mentionnons des problèmes d'optimisation combinatoire à objectifs non-linéaires (OWA, Choquet).
- b) l'étude de formulations étendues de taille polynomiale pour des polyèdres combinatoires : il s'agit de prolonger et si possible d'étendre à d'autres types de structures les résultats originaux obtenus récemment sur le polytope des arbres de Huffman.
- c) l'étude d'algorithmes exacts ou approchés de type primal-dual basés sur de nouvelles formulations ou relaxations linéaires des problèmes combinatoires impliquant des cycles, des chemins, des arbres. Les résultats obtenus seront appliqués à des variantes multicritères de divers problèmes combinatoires et au problème difficile de partitionnement de graphes.
- d) la poursuite de l'étude d'algorithmes efficaces et de cas particuliers polynomiaux pour une classe de problèmes d'optimisation robuste (les problèmes de programmation linéaire robuste avec incertitude sur les seconds membres) dont les applications industrielles sont particulièrement importantes comme nous l'avons montré (voir publication dans *Journal of Global Optimization* (2011), référence [33] de la partie bilan).

Evolution par rapport à la période précédente : sur la base de compétences reconnues depuis de nombreuses années dans le domaine des graphes, de l'optimisation continue et en nombres entiers, et des applications aux réseaux (logistique, Télécom, énergie, ...), deux évolutions notables viennent enrichir le champ d'investigation de l'équipe : l'une en direction de la problématique de la décision multicritères ; la seconde en direction de l'optimisation robuste, et en particulier pour la commande optimale robuste de systèmes dynamiques.

Originalité Les principales originalités des recherches menées par l'équipe MOS actuellement concernent :

- l'étude approfondie d'une classe importante de problèmes de programmation linéaire robuste, celle pour laquelle l'incertitude porte sur les seconds membres, cette étude a conduit à de nouveaux résultats de complexité ainsi qu'à la mise en évidence de nouveaux cas particuliers polynomiaux ; cette classe, jamais étudiée en tant que telle jusqu'ici, se rencontre dans de très nombreux domaines d'application ;
- l'identification d'une nouvelle hiérarchie de relaxations pour les problèmes de programmation mixte permettant une résolution plus efficace par rapport aux techniques existantes (RLT) grâce à l'exploitation de propriétés de décomposabilité ;
- la mise en évidence de nouveaux algorithmes approchés avec garanties de performances, fondés sur des approches primales-duales, en particulier pour divers problèmes de chemins et de cycles dans les graphes.

Objectifs de résultats à 5 ans Une quinzaine d'articles originaux dans des revues scientifiques internationales sur les différents sujets répertoriés dans la description du projet scientifique ci-dessous, présentations afférentes dans les congrès, échanges de chercheurs, encadrement de plusieurs thèses, etc.

Outre ces trois thèmes qui touchent aux aspects fondamentaux du domaine, nous souhaitons poursuivre et développer notre activité autour des systèmes de décision, en particulier sur l'acquisition des préférences utilisateurs (élicitation par questionnaire, apprentissage), sur l'apport des modèles graphiques pour la génération de recommandations (réseaux GAI, réseaux Bayésiens, Graphes contextuels), sur l'apport de l'analyse multicritère pour la génération de recommandations, enfin sur l'explication des décisions et l'analyse de leur robustesse. Notons que l'arrivée de Paolo Viappiani devrait permettre de renforcer encore cette activité sur les systèmes de décision et développer les aspects acquisition de préférences en décision multi-attribut comme en décision dans l'incertain. Nous prévoyons notamment de travailler sur l'élicitation d'utilités non-additives (e.g. intégrale de Choquet, thème 1) et sur l'apprentissage de modèles graphiques (thème 2) pour les systèmes de recommandation.

5 Analyse SWOT

Points forts : Double culture IA/RO de l'équipe et publications dans les deux communautés (bonne présence dans supports sélectifs), développement de sujets nouveaux au carrefour de la théorie de la décision et de l'optimisation combinatoire, visibilité nationale et internationale, fort potentiel applicatif des problématiques abordées, bonne implication dans les projets académiques et industriels, développement de synergies locales (équipes RO et SMA du département DESIR au LIP6 ; groupe "Apprentissage pour la Commande et la Décision en Robotique" à l'ISIR). Arrivée d'un CR1 pour renforcer les aspects élicitation de préférence et systèmes de recommandation.

Points faibles : Difficulté de maintenir un taux de présence et un flux de productions régulier dans différentes communautés simultanément (IA, Décision, RO). Par ailleurs, les productions de l'équipe ont un fort potentiel applicatif (visible par nos applications dans des domaines variés) mais sont loin d'avoir toutes rencontrées leurs utilisateurs potentiels. Nous souhaiterions augmenter encore le transfert industriel et notre présence auprès des industriels, sans toutefois diminuer notre niveau d'activité sur le plan académique.

Risques liés au contexte : Forte concurrence d'équipes voire de laboratoires entièrement dédiés à nos problématiques en France ou à l'étranger.

Possibilités liées au contexte : Au niveau local, concentration de compétences IA et Optimisation dans le département DESIR mais aussi dans d'autres équipes du LIP6. Existence d'un projet de thème transverse "optimisation" qui rassemble les équipes de plusieurs départements du LIP6. Vivier d'étudiants en informatique à l'UPMC avec un profil IA/RO et la possibilité de se spécialiser en décision (master IA et Décision, futur master ANDROIDE). Plus généralement, on bénéficie d'une concentration de chercheurs en AD/RO rassemblés sur Paris (LAMSADE à P9, LIP6 P6, CERMSEM P1) en bonne interaction, développement de la ROADEF, lancement récent du programme Gaspard Monge pour l'optimisation, la recherche opérationnelle et la gestion de l'énergie (EDF, fondation mathématique Jacques Hadamard), création récente du GDR International ALGODEC (théorie de la décision algorithmique) dans lequel nous sommes fortement impliqués.

6 Mise en oeuvre

Valorisation et recherche de moyens : L'équipe est largement impliquée dans les projets ANR nationaux en IA, AD ou RO et participe à des projets Européens. Nous envisageons de poursuivre notre activité au sein de projets académiques nationaux et internationaux, mais aussi de maintenir un flux régulier de partenariats industriels et d'augmenter le taux de thèses financées sur projet. Concernant les projets académiques, outre la poursuite des projets ANR en cours, nous démarrons un projet sur le thème "Morphologie mathématique et décision" avec des collègues du LTCI (Telecom ParisTech) et de l'équipe PEQUAN du LIP6 sur un financement Télécom ParisTech, dans le cadre des projets Futur et Ruptures. Concernant les projets industriels, mentionnons la soumission récente du projet ANR SIMU en collaboration avec l'IRSN qui vise à développer une plateforme de simulation d'accidents nucléaires graves. Notre contribution porte sur l'analyse du risque et la détermination des scénarios de catastrophe les plus probables. On peut également mentionner le projet ASTRID, un projet DGA / ANR en collaboration avec l'équipe Péquan qui vise à proposer des algorithmes de détection d'événements inhabituels dans des séquences vidéos (dépose d'un bagage sur le quai d'une gare, etc). Le projet vient d'être accepté. Enfin un projet OpenInsight en collaboration avec l'équipe MALIRE qui vise à proposer des outils fondés sur l'analyse de grosses masses de données pour aider les décideurs métiers à mieux gérer leur business sur le web.

Diffusion des résultats (stratégie de publication) : L'équipe développe des travaux au carrefour de la recherche opérationnelle et de l'intelligence artificielle et peut, selon le type de résultats obtenus, publier dans les deux communautés. La stratégie de l'équipe vise à rester présente aussi bien dans la communauté RO que IA. En particulier nous publions dans les revues intercommunautaires des deux communautés. Pour les conférences internationales, nous privilégions, quand le sujet s'y prête, des publications dans les conférences d'IA qui offrent des supports plus sélectifs tels que les actes des conférences IJCAI, AAAI, UAI, ECAI, AAMAS dans lesquels nous souhaitons continuer de publier. Néanmoins, nous sommes également présents dans les conférences internationales et nationales de RO pour participer à la vie de la communauté.

Stratégie locale / nationale/ Internationale : Au niveau de l'UPMC, l'équipe décision prévoit de poursuivre et développer ses interactions locales, en particulier avec les équipes RO (optimisation combinatoire, théorie des jeux algorithmique) et SMA (décision multi-agents, mécanisme design) du département DESIR au LIP6, avec le groupe "Apprentissage pour la commande et la décision en robotique" à l'ISIR (planification dans l'incertain) et avec l'équipe PEGUAN (modèles graphique et filtre particulaire). Par ailleurs, l'équipe joue un rôle moteur dans le GDR International 'AlgoDec' dédié à la théorie de la décision algorithmique. Nous participons activement au développement de cette thématique et à l'animation de manifestations internationales telle que la conférence ADT (Algorithmic Decision Theory) initiée en 2009, en interaction avec nos collègues du LAMSADE et du CRIL mais aussi du DIMACS (USA). Par ailleurs, l'équipe décision est impliquée dans le programme de travail du LABEX SMART notamment sur l'axe "Digital and Human Reasoning" avec les aspects élicitation de préférences, représentation de préférences, décision et planification multi-objectif, raisonnement et décision, systèmes de recommandation, choix social computationnel. Enfin l'équipe décision porte le projet ANR GUEPARD (collaboration avec le LAMSADE et le LINA) et prévoit d'organiser en 2013 un workshop de synthèse sur le thème de l'optimisation combinatoire multi-objectifs.

Implication future dans l'enseignement : Les travaux de l'équipe décision sont au coeur de la spécialité de master ANDROIDE proposée pour la nouvelle maquette du Master d'informatique de l'UPMC. Cette spécialité ANDROIDE héritière de l'actuelle spécialité IAD (IA et Décision) vise à former des étudiants en Intelligence Artificielle, Décision et Recherche Opérationnelle et comportera des enseignements principalement dispensés par les enseignants chercheurs du département DESIR sur les thèmes "AgeNts Distribués, Robotique, Recherche Opérationnelle, Interaction, Décision". Il s'agit là d'une des rares spécialités de Master qui offre en France l'opportunité aux étudiants de se former simultanément aux techniques de l'IA et de la RO aussi bien pour l'aide à la décision humaine que pour la décision automatique et les systèmes de décision. La plupart des membres de l'équipe Décision sont responsables d'au moins une UE de cette formation et Christophe Gonzales est porteur de cette spécialité ANDROIDE avec Nicolas Maudet (équipe SMA du département DESIR).



Vague D : Campagne d'évaluation 2012-2013

4.4 Projet 2014–2018 - Equipe SMA

1 Membres permanent de l'équipe

Nom	Prénom	Status	Etablissement
Beynier	Aurélie	MC	UPMC
Corruble	Vincent	MC	UPMC
El Fallah Seghrouchni	Amal	PU1	UPMC
Guessoum	Zahia	MC	IUT Reims
Kant	Jean-Daniel	MC	UPMC
Maudet	Nicolas	PU2	UPMC

2 Evolution de l'équipe

L'équipe a connu 2 départs de MC habilités à diriger les recherches. Nicolas Sabouret, a été nommé Professeur à l'Université Paris 11 - Orsay (Septembre 2012) et Samir Aknine a été nommé Professeur à l'Université Claude Bernard Lyon 1 (Septembre 2009). L'équipe prévoit de ré-intégrer Jean-Pierre Briot, Directeur de Recherches au CNRS, actuellement associé à l'équipe SMA et directeur du bureau du CNRS au Brésil. François Pachet, chercheur Senior à Sony CSL souhaite rejoindre l'équipe SMA (à partir de Septembre 2012) en tant qu'associé à l'occasion de son ERC. L'équipe prévoit de recruter un MC et un Professeur des Universités sur les thématiques émergentes de l'équipe du fait de leur importance et de la demande pressante en termes d'encadrement et d'opportunités au niveau Européen et industriel.

3 Caractérisation de la recherche

Positionnement recherche fondamentale et recherche appliquée

Nos recherches sont à la fois fondamentales et appliquées comme peuvent en témoigner nos publications dans les journaux et conférences majeurs du domaine SMA (ajouter nombre) et également par nos collaborations industrielles aussi bien dans le cadre de contrats de recherche que de thèses financées par l'industrie (7 thèses CIFRE et 8 financées par des contrats)

4 Projet scientifique

– Originalité et positionnement nat/ Int.

L'équipe SMA s'intéresse aux agents cognitifs comme paradigme de conception et de programmation en vue d'apporter aisance et intelligibilité aux concepteurs des systèmes intelligents et distribués. Nous développons ainsi 1) des modèles de conception et/ou de simulation multi-agents dans lesquels les agents sont autonomes (ex. pour décider, planifier, apprendre, etc.), capables d'action et d'interaction avec d'autres agents (ex. se coordonner, négocier, etc.) et l'environnement (ex. percevoir et réagir) 2) des langages de programmation orientée agents et des plates-formes support pour la programmation et le déploiement de SMA ouverts vers des utilisateurs et/ou d'autres systèmes. Ces modèles et langages

nous permettent de réaliser des applications de portée industrielle dans divers domaines comme l'intelligence ambiante, les systèmes de missions, les jeux vidéo, la simulation de systèmes complexes, etc. Pour toutes ces applications, nous mettons au centre de nos conceptions l'autonomie des agents, leur capacité d'action individuelle, mais aussi collective, avec des modèles sophistiqués de coordination et de prise de décision distribuée.

Dans le domaine des SMA, des progrès significatifs ont été observés ces dernières années sur des sujets comme l'autonomie et l'adaptation, l'ingénierie, la tolérance aux fautes et la fiabilité, l'informatique affective et la prise en compte des émotions, la simulation multiagents, ou encore les capacités de prise de décision collective.

- **Evolutions par rapport à la période précédente /Prise de risque** On peut organiser les nombreux verrous qui animent les recherches actuelles sur les SMA selon trois points de vue : du système lui-même, de l'utilisateur final ou observateur, et du concepteur.

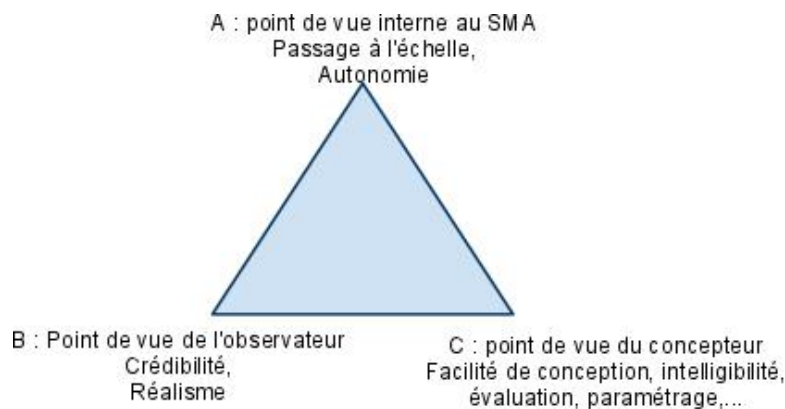


FIGURE 4.1 – Trois points de vue sur les verrous des SMA

La difficulté, et donc le risque associé, vient de ce que beaucoup de situations réalistes combinent 2 ou 3 de ces verrous simultanément. Cette difficulté est en elle-même un défi scientifique au coeur de la problématique de l'équipe SMA.

Au vu des compétences de l'équipe, des évolutions thématiques multi-agents et du contexte scientifique, industriel et sociétal, l'équipe souhaite mettre en avant deux axes majeurs de recherche pour les 5 à 10 années à venir.

- **Positionnement par rapport aux thèmes transverses et de département** Plusieurs de thématiques mentionnées ici trouvent des échos ou des prolongements dans le cadre du département ou de thèmes transverses. En ce qui concerne par exemple la problématique du passage à l'échelle, on peut noter que les modèles étudiés dans la communauté pour analyser les situations d'interactions stratégiques impliquant un grand nombre d'agents, les *jeux graphiques*, mettent en jeu des principes très proches de ceux des modèles graphiques pour le raisonnement et la décision (réseaux Bayésiens, réseaux de préférence, etc.). Nous avons engagé sur cet axe un groupe de travail exploratoire impliquant outre notre équipe, celle de Décision et celle de RO. Le thème transverse "Social Computing" recoupe de nombreuses problématiques au coeur de nos recherches (modélisation de systèmes complexes, dynamiques et organisation au sein des communautés...). Ce thème offre l'opportunité de consolider des collaborations avec des spécialistes de la dynamique des réseaux ou de l'analyse de données sociales. L'équipe est également présente dans le thème transverse "Optimisation", essentiellement par le biais des problématiques de décision distribuée dans lesquels le résultat, la communication, ou le stockage peuvent être des critères à optimiser. Par ailleurs, l'équipe SMA co-anime avec l'équipe NPA du LIP6 un axe transverse intitulé "Interactions Intelligentes et Mobiles" qui regroupe 13 équipes du LIP6 de 4 départements scientifiques. Cet axe a pour objectif de traiter de façon intégrée 2 niveaux habituellement étudiés séparément : l'infrastructure de communication (objets communicants) et l'intelligence ambiante qui constituent enjeu majeur pour les années à venir, tant sur le plan scientifique que du point de vue des applications industrielles, commerciales et sociétales. Relativement à cet axe, l'équipe des SMA s'intéresse à la dynamique des interactions, la modélisation et simulation multi-échelle, l'algorithmique répartie et les protocoles de coordination, la vérification et l'analyse comportementale en contexte et la distribution de l'intelligence

sur l'infrastructure.

4.1 Interactions Virtuel - Réel : les SMA de la simulation à l'intelligence ambiante

Description : Deux des trois thèmes décrits dans le bilan de l'équipe SMA, nommément le thème *Modélisation et Simulation Multi-Agents* et le thème *Ingénierie des SMA*, ont vocation à converger dans les années à venir. L'équipe a en effet repéré entre ces deux thématiques de vraies opportunités pour de fortes synergies et fertilisations croisées, ainsi que des applications porteuses. Cette idée est fortement inspirée de travaux maintenant classiques sur la notion de réalité mixte (MR), illustrée ci-dessous à partir des travaux de Milgram et Kishino¹ (1994), qui introduit un continuum entre les notions d'environnements virtuels et réels (Figure 4.2).

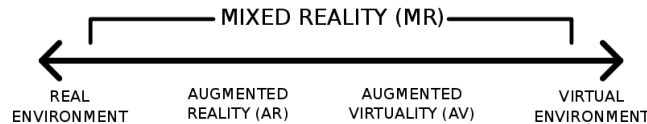


FIGURE 4.2 – Continuum : réalité mixte

Ces travaux classiques se focalisaient sur la représentation visuelle des environnements, notre démarche vise à les transposer aux domaines des systèmes multi-agents où les éléments clés de la réalité physique et virtuelle ne sont plus les éléments visuels mais des agents autonomes capables d'interagir entre eux et avec l'environnement (Figure 4.3).

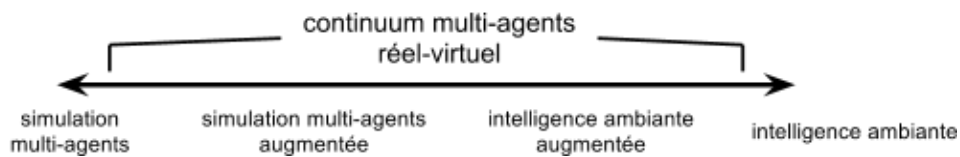


FIGURE 4.3 – Continuum : multi-agents

En simulation multi-agents, ainsi que dans de nombreux jeux et jeux sérieux, on étudie des systèmes souvent complexes à travers leur modélisation virtuelle. C'est la congruence entre le modèle (virtuel) et le système (réel) simulé qui donne à l'exercice son intérêt (par exemple, meilleure compréhension du système ou capacité de prédiction sur les états futurs pour la simulation ou le jeu sérieux, meilleure immersion, facteur d'engagement et d'amusement pour le jeu vidéo). La simulation est liée au monde réel à travers le processus de modélisation, mais aussi, et c'est une thématique de recherche importante, à travers le processus de calibration qui renseigne les paramètres de la simulation grâce à des données issues du monde réel. Souvent ce processus de calibration est une étape initiale pour la simulation, mais dans un nombre grandissant de domaines et d'applications, la cohérence entre réel et virtuel est établie ou maintenue de manière répétée voire continue en cours de simulation. Citons quelques exemples :

- Les Jeux ubiquitaires (jeux urbains). L'activité dans le monde réel (par exemple la position dans la ville pour un jeu urbain) a un impact direct sur l'évolution du monde virtuel du jeu ;
- La simulation et le jeu sérieux : des éléments de la réalité (exemple simple : la situation météo au temps t, l'encombrement réel d'une route, les cours de la bourse, les actions d'un décideur) sont pris en compte dans la simulation (simulateur de vol, simulation de ville, simulation économique, jeu sérieux) ;
- la simulation de systèmes complexes humains (ex : ville, économie du travail, organisation de l'entreprise, etc.) vise à produire des outils d'aide à la décision pour des politiques publiques, aide dont la pertinence dépend fortement de la capacité à reproduire précisément des faits réels observés et à les expliquer.

De plus, si le monde réel est lui-même peuplé de dispositifs intelligents capables d'autonomie et d'interactions entre eux et avec les êtres humains, on obtient des systèmes d'intelligence ambiante, une des

1. P. Milgram and A. F. Kishino, *Taxonomy of Mixed Reality Visual Displays* IEICE Transactions on Information and Systems, E77-D(12), pp. 1321-1329, 1994.

thématiques importantes pour l'équipe SMA. Ces systèmes peuvent eux-mêmes tirer parti de la simulation soit pour aider à leur conception (évaluation des concepts avant déploiement en particulier), soit en cours d'utilisation pour accéder à des informations qui ne sont pas accessibles dans le monde réel.

Le potentiel pour les SMA aux deux niveaux (réels, virtuels) et pour leurs interactions est alors énorme, et peut bien s'illustrer à travers la notion applicative de ville intelligente (*smart city*) où les problématiques de recherches et les débouchés potentiels sont nombreux, et peuvent bénéficier de l'expérience acquise par l'équipe sur la simulation de villes (projet TerraDynamica) et dans le domaine de l'Intelligence Ambiante. Comme les travaux sur les *smart cities*, en particulier si l'on vise des systèmes déployables pour des expérimentations en grandeur réelle, s'inscrivent plutôt sur le long terme, nos travaux actuels sur des salles intelligentes (Syamisen, smart room, projet AoDAI) ou entamés sur le concept de campus intelligent, constituent une preuve de concept et peuvent donner lieu à des systèmes déployés à court et moyen termes. Par ailleurs, cette nouvelle thématique (section 4.1) permettra aussi à l'équipe de poursuivre ses efforts pour s'attaquer à des systèmes de complexité croissante, appréhender le passage à l'échelle de nos modèles (ex. ville, économie du travail). Nous approfondirons en particulier nos travaux sur la modélisation multi-niveaux afin d'être en mesure d'étudier un système complexe à des niveaux d'abstractions différents en permettant le passage automatique et dynamique d'un niveau à l'autre.

Originalité et positionnement dans le contexte local/ national/ international : L'équipe SMA a une expertise significative sur les domaines contribuant à couvrir le continuum du réel au virtuel (4.3) allant de la conception à la mise en oeuvre. Elle est donc très bien placée aux niveaux national et international pour explorer ce thème qui nous semble particulièrement riche et novateur du point de vue de la recherche et des applications. Par ailleurs, l'équipe bénéficie sur cette thématique d'un contexte favorable au sein du LIP6 (voir prospective AMI&OC).

4.2 Coordination et Consensus dans les SMA ouverts

Description : Cet axe couvre les modèles de coordination multi-agents et la recherche de consensus entre agents. Nous nous intéressons à la validation des solutions garantissant certaines propriétés comme la convergence, une certaine optimalité, en dépit de contraintes des modèles, qu'ils s'adressent à des agents coopératifs ou compétitifs. Une des évolutions majeure de cet axe portera sur la prise en compte de contraintes spécifiques liés aux systèmes complexes du monde réel que nous étudions (voir axe 1).

L'**ouverture du système** (des agents entrent et sortent du SMA) peut remettre en cause la validité des décisions ou des consensus précédemment établis. Cette problématique est très peu abordée dans les travaux de recherche actuels mais sa prise en compte est indispensable dans les systèmes réels. Par exemple, les modèles actuels traitant de décisions distribuées sous incertitude considèrent tous des systèmes d'agents fixes. L'ouverture du système pose la question de la mise à jour des connaissances des agents sur le système, de la validité des décisions ou des consensus qui ont été établis, de la replanification,... De même, les modèles qui fondent les études de choix social, issus de la théorie du vote, se basent pour l'essentiel sur l'hypothèse d'un ensemble fixe et connu d'agents et d'options disponibles. Dans le courant du *choix social computationnel*, ces hypothèses sont relâchées et les aspects dynamiques commencent à être étudiés : connaissance partielle des ensembles de votants ou de candidats, solutions robustes à des aléas. Nous participons largement à cet élan. Cet axe rejoint, par certains aspects, les problématiques d'*algorithmique en ligne*, qui s'intéressent à un type particulier de dynamique (voir la prospective Optimisation).

La **mobilité des agents** au sein de l'environnement peut impacter la topologie du SMA. Une des conséquences majeures dont il faut tenir compte est l'impact sur les communications : les communications entre certains agents peuvent devenir impossibles ou non fiables, les coûts de communication sont impactés. Cela pose des questions cruciales de meilleures politiques de diffusion de l'information, de mise en place de protocoles adaptés. Formellement, les problèmes peuvent être étudiés sous l'angle de la complexité de communication, cela permet d'apprécier la qualité des protocoles proposés. Ces outils sont cependant peu adaptés (bornes peu pertinentes) à l'étude de topologies dynamiques envisagées ici. De nouveaux outils seront à mettre en place dans cette optique. De même, des compromis peuvent être à chercher entre qualité des solutions trouvées et quantité de communication requise.

Nous souhaitons également aborder la question de l'**implémentabilité de la coordination et du consen-**

sus dans/pour ce type de SMA. Dans les systèmes où les agents ne peuvent pas être supposés coopératifs, il est important de chercher à se prémunir des comportements de manipulation. Ainsi, un agent peut avoir intérêt à annoncer des préférences qui ne correspondent pas à la réalité afin d'obtenir une solution qui lui soit favorable. Ces questions, largement étudiées dans le champ de la conception de mécanismes (mechanism design), en particulier de mécanismes à véricité garantie, se déclinent naturellement dans le cadre multiagents, avec les contraintes déjà citées.

Le très grand nombre d'agents reste un paramètre crucial pour bon nombre de modèles qui nécessitent d'être repensés quand ce nombre devient très élevé. Cette question rejoint celle du passage à l'échelle déjà mentionnée dans l'axe « Interactions Virtuel-Réel ».

Originalité : L'intérêt de considérer ces hypothèses est de proposer des mécanismes réalistes et viables pour les SMA complexes que nous étudions (voir le thème 1 - et la prospective du LIP6 AmIOC). L'objectif est de proposer des méthodes de coordination et de recherche de consensus qui tiennent mieux compte de la complexité des situations dans des applications comme l'intelligence ambiante, la robotique mobile.

Positionnement dans le contexte local/ national/ international : L'axe de recherche sur le choix social computationnel s'inscrit dans un mouvement international important, qui a vu cette thématique émerger et se structurer à l'échelle européenne et mondiale. Nicolas Maudet est membre de l'Information and Communication Technologies COST Action « Computational Social Choice »(2012-2016). Par ailleurs, l'équipe bénéficie sur cette thématique d'un contexte favorable au sein du LIP6 (voir prospective Optimisation).

5 Analyse SWOT

Points forts : – Nos recherches vont du fondamental à l'expérimental (solutions complètes, compétences et expériences reconnues dans les domaines concernées par ce projet : langages d'agents, simulation, coordination, AOSE) ;

- Ouverture à la pluri-disciplinarité et aux problèmes de société ;
- Synergies avec des équipes du LIP6 (Prospective du LIP6, AMI, Optimisation, Social Computing).

Points faibles : Nécessité d'obtenir rapidement les moyens nécessaires : infrastructures, moyens humains (personnels, usagers) et matériels.

Risques liés au contexte : Concurrence au niveau national (ex. ADREAM) et international (beaucoup de projets de type Smart cities, smart Planet d'IBM..).

Possibilités liées au contexte : – Domaine de Recherche en plein essor ;

- Ecosystème favorable : Ile de France ; Pôles de compétitivité comme Cap Digital et Advancity, Labex SMART.
- L'intelligence ambiante est un des thème prospectif du CNRS ; Deux thèmes transverses en discussion au sein du GDR III (Intelligence ambiante et Systèmes Complexes).
- Vaste domaine d'applications ;
- Divers appels à projets (ex. Projets Européens. Exemple : 7ème PCRD - Smart Cities) ;
- Thématiques qui intéressent les industriels.

6 Mise en œuvre

Doit répondre en partie à l'analyse swot et aux faiblesses/ risques.

Valorisation et recherche de moyens : Partenariat, montage de projets

Diffusion des résultats (stratégie de publication) : Conférence de référence vs. workshops

Stratégie locale / nationale/ Internationale : UPMC-Region / LABEX - Animation communauté / Actions internationales : projets, réseaux

Implication future dans l'enseignement : Lien avec la nouvelle maquette du L et M, Echanges internationaux



Vague D : Campagne d'évaluation 2012-2013

4.5 Projet 2014–2018 - Equipe MOCAH

1 Membres permanents de l'équipe

Nom	Prénom	Status	Etablissement
Carron	Thibault	MC	Univ. Savoie
Delozanne	Elisabeth	MC	UPMC
Labat	Jean-Marc	Pr	UPMC
Yessad	Amel	MC	UPMC

2 Evolution de l'équipe

L'équipe va recruter un MC en 2013.

3 Caractérisation de la recherche

Le domaine des Environnements Informatiques pour l'Apprentissage Humain – EIAH a une caractéristique très spécifique : Il est situé au croisement des recherches fondamentales et finalisées et ces recherches fondamentales ont un fort caractère pluridisciplinaire dans la mesure où les modélisations informatiques sont élaborées en partie pour tester et si possible valider des modèles issus des sciences humaines et sociales. Les recherches en EIAH relèvent donc d'un double enjeu :

- co-élaborer avec des chercheurs en sciences humaines des modélisations exécutables dans un processus itératif de révision et d'expérimentation des modèles de sciences humaines et informatiques
- concevoir et produire des ressources, des outils et des environnements numériques qui soient utilisables et utiles pour l'enseignement et la formation

C'est un enjeu de société fondamental mais ce passage à l'échelle de l'utilisation du numérique est encore loin d'être accompli tant les problèmes, qu'ils soient de nature cognitive, technique ou organisationnelle, sont loin d'être résolus.

MOCAH participe à ce double mouvement en menant d'une part des recherches fondamentales sur la modélisation informatique de la cognition humaine à travers des projets pluridisciplinaires, en particulier en mathématiques, et d'autre part des recherches finalisées sur la conception d'outils auteurs, soit pour la production de serious games par des informaticiens (bourse Cifre Capdevila), soit pour la production ou l'adaptation de ressources numériques par des enseignants (chaîne éditoriale Exo+, thèse de B. Marne sur un outil auteur pour adapter des serious games existants).

4 Projet scientifique

La première priorité est de mettre l'essentiel de nos capacités de recherche sur les problématiques liées au développement des serious games. Sur ce sujet, nous voyons de nombreux thèmes de recherche se dessiner (voir la liste ci-dessous) qui sont innovants et dont certains comportent une part de risque.

En deuxième lieu, nous souhaitons mettre nos compétences sur le diagnostic cognitif et en modélisation de l'apprenant au service des thèmes transverses du Lip6 dans lesquels une modélisation de l'utilisateur (donc,

au-delà de celle de l'apprenant) est nécessaire, i.e. sur le thème de la E-santé, les réseaux sociaux (qui ont un impact sur la e-éducation) et sur le thème de l'intelligence ambiante, thème mené par l'équipe SMA du département DESIR.

Enfin, nous souhaitons continuer à valoriser nos recherches, en intégrant d'avantage nos logiciels dans les plateformes de nos partenaires industriels ou académiques. Cette phase de transfert nous permet d'améliorer nos logiciels grâce aux retours d'utilisateurs réels et de mieux conduire nos recherches.

4.1 Thème 1 : serious gaming

Sur ce thème, que nous avons été parmi les premiers à aborder en France en 2008, l'équipe souhaite s'attaquer à la question des stratégies d'apprentissage, au verrou qu'est l'adaptation à l'apprenant et sur la conception d'un DSML dédié à la validation de l'adéquation du serious game aux objectifs en terme d'acquisition des connaissances donnés par l'expert pédagogique. Nous voulons élaborer de nouvelles stratégies d'apprentissage qui soient basées sur des serious games multi-joueurs dans lesquels les joueurs sont amenés parfois à être en compétition, parfois à collaborer (la "coopétition").

Pour prendre en compte les modes de vie des jeunes, de plus en plus mobiles et connectés, y compris avec la TV connectée, nous voulons concevoir des stratégies d'apprentissage ubiquitaire, reposant sur un mode d'apprentissage mixte, distanciel et présentiel, dans le prolongement des stratégies multi joueurs. Proposer des modèles pour adapter dynamiquement les serious games aux comportements des joueurs.

Construire un DSML (Domain-Specific Modeling Language) dédié à la vérification formelle des spécifications des serious games. Ce langage permettrait aux équipes de conception d'exprimer formellement les propriétés que doit vérifier le serious game. Nécessaire dans le cas mono-joueur, ces vérifications deviennent indispensables mais relativement complexes à mettre en oeuvre dans le cas multi-joueurs.

Même si, aujourd'hui, de nombreuses équipes en France (Lyon, Grenoble, Montpellier, Valenciennes, Toulouse) et dans le monde (London Knowledge Lab, InGaMe Lab, SGI de Coventry University, Texas A & M University, e-UCM e-learning research group at Universidad Complutense de Madrid, etc.) travaillent sur les serious games, les thèmes listés ne sont guère abordés par les autres équipes, en particulier les 2 derniers verrous où nous pourrions bénéficier de collaborations au sein du Lip6 : Sur le 3ième verrou, nous souhaitons tester des modèles de décision multicritères et dans l'incertain tels que proposés par l'équipe DECISION. Sur le 4ième verrou, qui est un objectif à 5 ans, nous espérons pouvoir y obtenir des résultats grâce à une collaboration déjà commencée avec l'équipe MOVE.

4.2 Thème 2 : modélisation de l'utilisateur

Un modèle utilisateur est une représentation explicite des caractéristiques des utilisateurs interagissant avec un système. Ce n'est jamais une modélisation générale mais une modélisation adaptée au contexte. Cette modélisation peut être faite sur les plans cognitif, affectif, comportemental et intentionnel. Dans le cadre des projets transversaux, les compétences du LIP6 permettent d'attaquer les verrous suivants : Suivi de la personne, nouveaux types d'information dans le modèle de la personne et algorithmes de traitement adaptés (ex. modélisation par réseaux bayésiens), construction de modèles issus de l'informatique affective pour des agents artificiels en interaction avec la personne (agents assistants et compagnons), conception de capteurs intelligents qui permettent la caractérisation de la personne et de son environnement, en particulier de son état physique mais aussi émotionnel. Dans ce contexte, MOCAH continuera le travail sur le diagnostic cognitif à partir de l'analyse des traces mais aussi sur la conception d'architectures cognitives de type SOAR ou ACT R avec une approche pluridisciplinaire ainsi que sur une architecture générique pour développer des modèles d'utilisateurs, notamment pour les applications de e-santé.

Sur le plan national, c'est essentiellement l'équipe METAH du LIG qui aborde la modélisation de l'utilisateur. En revanche, il y a une très forte communauté internationale, structurée autour de l'association "User Modeling" avec une revue et une conférence de niveau A+ et de nombreux laboratoires parmi les plus prestigieux (entre autres Carnegie Mellon avec J-R Anderson et K. R. Koedinger) qui travaillent sur ces problématiques.

4.3 Thème 3 : Intégration de la chaîne éditoriale Exo+

Nous avons participé à 2 soumissions à l'appel "E-éducation 2" des emprunts d'avenir afin que la chaîne éditoriale Exo+ déjà utilisée par les éditions Pole soit intégrée dans un ENT. Même si ces 2 réponses n'étaient

pas retenues, nous avons l'objectif de continuer avec différents partenaires industriels et académiques (y compris en interne à l'UPMC) de rendre complètement opérationnelle cette chaîne qui permet de créer des exercices mais aussi des tests, thème sur lequel nous continuons en parallèle de mener des recherches pour les améliorer (création de tests adaptatifs pour l'évaluation ou la remédiation). Nous sommes les seuls à travailler en France sur ce thème. Au niveau international, les principaux concurrents sont l'université du Massachusetts (T. Murray) et celle de Floride (V. Shute). Par ailleurs, nous sommes très intégrés et très reconnus dans le consortium IMS QTI.

5 Analyse SWOT

Points forts : Sur nos domaines d'expertise (outil auteur selon le standard IMS QTI, diagnostic cognitif en mathématiques, serious gaming), l'équipe est tout à fait reconnue. Elle a obtenu 8 projets lors des différents appels à projet, et nous avons mené à terme 4 réalisations logicielles d'envergure, avec une diffusion et une utilisation réelle.

Points faibles : L'effectif de l'équipe, en diminution depuis 2010, restera faible malgré un recrutement et deux nouvelles arrivées espérées. Il est difficile d'être visible au plan international dans un domaine où l'expérimentation est importante.

Possibilités liées au contexte : En Ile de France, de nombreux chercheurs participent activement au domaine des EIAH au sein de différents laboratoires (LDAR en didactique, Paragraphe en psychologie cognitive, EDA en sciences de l'éducation, LTCI en sciences du langage, MSH en sciences de l'information, et Lip6 en informatique). D'autre part, une communauté de domaine du pôle de compétitivité CAP DIGITAL porte sur ce thème avec de nombreuses entreprises très actives (Editis, Hachette, Microsoft, Ktm-Advance, Knowmore, Strass Production, Wizarbox, etc.).

Risques liés au contexte : Il est essentiel d'avoir des collaborations, non seulement avec des partenaires industriels mais aussi académiques pour pouvoir développer des projets d'envergure internationale, en particulier pour participer à des projets européens.

6 Mise en oeuvre

Nous espérons dans l'avenir proposer une structure fédérative avec les équipes mentionnées ci-dessus, dont certaines participaient déjà au PPF AIDA. Cela nous donnerait une réelle visibilité au niveau national et international.

En ce qui concerne les publications en revue, nous veillerons à présenter des articles plus nombreux se focalisant sur un résultat limité plutôt que publier des articles de 40 pages, ce qui a été souvent notre cas.

Nous espérons dans l'avenir proposer une structure fédérative avec les équipes mentionnées ci-dessus, dont certaines participaient déjà au PPF AIDA. Cela donnerait une réelle visibilité au niveau national et international. En ce qui concerne les publications en revue, nous veillerons à présenter des articles plus nombreux se focalisant sur un résultat limité plutôt que publier des articles de 40 pages.

Valorisation et recherche de moyens : Nous avons une préoccupation constante de diffusion de nos outils afin qu'ils soient utilisés en situation réelle, ce qui en retour nous permet d'y apporter des améliorations. C'est le cas pour le logiciel Combien ?, pour notre chaîne éditoriale Exo+, pour le logiciel lingot avec l'association Sesamath et les outils de développement des serious games réalisés dans le cadre d'une thèse cife.

L'équipe a obtenu de nombreux projets lors de ces 5 dernières années et souhaite poursuivre dans cette voie qui lui permet d'avoir en permanence entre 1 et 2 post-docs et parfois un développeur, ressources humaines importantes dans notre domaine pour réaliser des logiciels que l'on puisse expérimenter en situation réelle. Nous avons participé en 2012 à 2 appels à projets européens (un projet Interreg IVB North-West Europe et un projet TEMPUS Europe de l'Est) pour lesquels nous n'avons pas encore la réponse.

Diffusion des résultats (stratégie de publication) : La stratégie de l'équipe est de ne pas se disperser en multipliant les publications dans des conférences de 2ème ordre ou des workshops, sauf lors des débuts de thèse, mais de se centrer sur les conférences et revues de niveau A.

Stratégie locale / nationale/ Internationale : Au niveau local, l'équipe souhaite continuer de participer à la communauté de domaine E-éducation de CAP DIGITAL afin de développer ses partenariats industriels et, sur le plan académique, l'équipe va poursuivre sa stratégie de regroupement pluridisciplinaire. Au niveau national, l'équipe est reconnue comme une des équipes importantes du domaine (malgré sa taille réduite) et est un élément moteur tant au sein de l'association scientifique du domaine (ATIEF) que dans l'organisation scientifique de la conférence francophone EIAH et dans la revue STICEF. Au niveau international, nous sommes membre de l'association TELEARC qui s'est créée après la fin du réseau d'excellence Kaleidoscope. Nous comptons poursuivre notre implication dans le consortium IMS QTI et nous entretenons des contacts réguliers avec les équipes européennes travaillant sur les serious games, ce qui a fait que nous avons été sollicités pour participer à la soumission des 2 projets cités ci-dessus.

Implication future dans l'enseignement : La spécialité "Ingénierie de la Formation en Ligne", renommée "Ingénierie et Management de la Formation en Ligne" du master " Management des connaissances et Innovation" devrait se poursuivre avec une évolution importante en ce qui concerne le dispositif puisque nous avons l'objectif de l'ouvrir en apprentissage et en formation continue avec un dispositif par alternance afin de mieux répondre aux demandes de nos partenaires, qu'ils soient académiques comme Créteil ou industriels, notamment les sociétés de E-learning. Dans le cadre de la nouvelle spécialité, "AgeNts Distribués, Robotique, Recherche Operationnelle, Interaction, Decision" du master Informatique, nous avons proposé une coloration "Environnements Interactifs" (simulations, jeux vidéo, serious games), avec la responsabilité d'une UE de S3 sur les serious games, UE que nous proposons aussi dans le cadre d'un programme européen de type TEMPUS.

Chapitre 5

Projet scientifique - Département DAPA

Sommaire

5.1	Projet 2014 – 2018 - Département DAPA	88
5.2	Projet 2014–2018 - Equipe BD	91
5.3	Projet 2014–2018 - Equipe LFI	96
5.4	Projet 2014–2018 - Equipe MLIA	100
5.5	Projet 2014–2018 - Equipe ACASA	105



Vague D : Campagne d'évaluation 2012-2013

5.1 Projet 2014 – 2018 - Département DAPA

1 Présentation du département

Le département est actuellement constitué des équipes MALIRE, BD et ACASA (respectivement 19, 7 et 3 permanents). Il évoluera en 2013 vers une structure en 4 équipes, à la suite de la création des équipes LFI¹ (8 permanents) et MLIA² (11 permanents) issues de MALIRE, dont la synergie existante sera préservée, en particulier par la continuation de réunions communes dans le cadre du “Forum apprentissage”.

Au niveau des membres, on notera le départ de M.-R. Amini (09/2012) et l'arrivée de Stuart Russell (09/2012) pour deux ans dans MALIRE sur une chaire Blaise Pascal, ainsi que le départ de M. Roux en (10/2012) et l'arrivée de G. Bourgne (09/2012) dans ACASA.

2 Périmètre scientifique

Les recherches du département DAPA sont centrées autour de l'apprentissage automatique (statistique, symbolique, relationnel, dans l'incertain) et de la gestion de grandes masses de données hétérogènes, réparties et dynamiques (flux). Le projet pour les cinq prochaines années concerne des avancées théoriques sur la conception de nouvelles techniques et d'algorithmes d'apprentissage et de traitement de données pour l'accès à l'information dans un environnement complexe. Ces techniques doivent permettre le passage à l'échelle et le traitement de grandes masses d'information, évoluant dans le temps, éventuellement distribuées, structurées ou non. Ces travaux se déclinent selon divers axes. La gestion de données et la recherche d'information textuelle et multimédia, sur des flux sociaux évolutifs, nécessitant la fusion d'informations hétérogènes et la prise en compte de sa validité, est le premier d'entre eux. Le traitement de relations complexes entre données ou entre sources, par exemple sur le web ou dans les réseaux sociaux sera également un axe de recherche. Un élément important de la recherche d'information est la prise en compte d'émotions dans les données textuelles ou multimédias et les réseaux sociaux ; il est commun à un autre axe de recherche qui a émergé récemment dans DAPA, lié à la fouille de données physiologiques. Les applications sociétales de ces divers travaux vont de la e-santé à l'internet des objets, en passant par l'exploitation de médias sociaux, la préservation des données, le tracking d'objets dans des vidéos ou les humanités numériques.

3 Positionnement

Au LIP6, les travaux prévus concernent tout naturellement les axes de recherche transversaux de “Traitement distribué de données distribuées”, “Social computing”, “Ultra large scale” et bien sûr “e-santé”.

Sur le plan national, les travaux développés par DAPA sont centraux pour la communauté d'apprentissage, par exemple au LIG, à l'ENS Ulm, au LRI, à Telecom ParisTech, celle de traitement de l'incertain (LIRMM, LISTIC, IRISA) ou encore de bases de données (CNAM, université de Cergy, INRIA, LIRIS, IRIT, LINA). Sur un plan international, le département DAPA se situe au cœur des travaux internationaux sur l'apprentissage et la recherche d'information, par exemple traités dans les universités de Bristol, de Berkeley, ou du London College, entre autres. Ses succès dans les compétitions internationales, les conférences invitées de ses membres sur le

1. “Learning, Fuzzy and Intelligent systems”

2. “Machine Learning and Information Access”

sujet, ses nombreuses publications, sa participation aux réseaux MUSCLE et PASCAL, et ses responsabilités dans les congrès internationaux sont la manifestation de son implication dans la communauté.

En ce qui concerne les médias sociaux, la richesse des données et la complexité des traitements posent de nombreux défis qui sont traités par les grands acteurs du Web 2.0 (Yahoo!, Google, Facebook) et des nombreux laboratoires reconnus au niveau international (Max Planck Institut, EPFL, IBM Research Lab, université d'Oxford, Stanford). Dans le domaine de la e-santé, le département se place parmi les chercheurs travaillant sur l'assistance aux personnes âgées ou en situation de handicap, aux États-Unis (université du Missouri), en Corée (KAIST), aux Pays-Bas (université d'Amsterdam), au Japon (université de Tsukuba). La fouille de données issues de capteurs dans cette optique ou plus largement dans le domaine de la maison intelligente est pratiquée, par exemple, aux USA (Carnegie Mellon) ou en Espagne (ECSC ou université de Granada). Le domaine des humanités numériques est très actif internationalement. Il existe des équipes aux États-Unis (MIT – Media Lab par exemple), en Suisse, à l'EPFL, en Italie, en Belgique au Royaume-Uni, avec qui nous avons des contacts à travers le réseau OBVIL.

Dans ce paysage très compétitif, le département DAPA se distingue par une volonté de couvrir plusieurs thématiques essentielles et complémentaires (apprentissage, bases de données, recherche d'information, web sémantique) pour aborder et résoudre des problèmes complexes de traitement d'information à large échelle. Cette orientation lui permet également d'élaborer des projets interdisciplinaires avec des sciences humaines (sociologie, littérature) et la médecine.

4 Les défis scientifiques

Les défis du département sont orientés dans cinq directions :

- **Ultra large scale.** Il correspond à l'une des perspectives transversales du LIP6, sous la responsabilité de DAPA. Ce défi concerne des problématiques d'apprentissage et d'indexation et il est abordé au niveau du volume des données, du nombre de classes ou de traitements et de la distribution. Il est rencontré dans la recherche d'information dans tous les médias, la fouille de données issues de capteurs et également dans les travaux sur l'optimisation multi-requêtes continues, les stratégies de rafraîchissement de données du web (flux RSS, pages Web) et la cohérence de données distribuées.
- **Données distribuées.** Elles sont naturellement présentes dans les réseaux sociaux, en particulier, et sont également partie prenante dans les méthodes de traitement de données à large échelle étudiées dans LFI et MLIA. L'équipe BD est concernée par ce défi dans ses activités sur la propagation efficace d'événements utilisateurs dans des réseaux sociaux distribués, la construction de fédération d'archives et la gestion de données complexes distribuées. L'équipe ACASA est impliquée sur les aspects "privacy".
- **Traitement de données atomiquement distribuées.** Le traitement de données issues de mobiles ou de capteurs fait partie des problèmes à résoudre dans certaines applications de la e-santé ou de l'internet des objets, par exemple. Il pose des problèmes de fusion sur lesquels le département pourra utiliser ses compétences.
- **Dynamisme.** Les systèmes dynamiques sont présents à tous les niveaux de la recherche d'information, des sources aux contenus. La prise en compte de flux d'information continus nécessite des méthodes spécifiques et leur traitement est l'un des défis des cinq prochaines années. L'évolution temporelle des réseaux ou des opinions est un point clé auquel se confrontera le département, ainsi que les problèmes de cohérence posés par la dynamique des données distribuées et répliquées.
- **Sémantique / Extraction.** Les problèmes liés à l'annotation constituent un défi en raison de la grande masse des informations et de l'attente des utilisateurs en termes d'étiquettes subjectives ou de caractérisations de haut niveau. Ces problèmes seront abordés dans LFI, MLIA et BD. Du côté d'ACASA, il s'agit d'une annotation sémantique massive à la fois sur des corpus littéraires, dans le cadre d'OBVIL, et sur des corpus scientifiques. Le problème de la cotation des données fait aussi partie des défis que relèvera l'équipe LFI, en particulier dans les réseaux de sources, et l'équipe BD s'intéressera au problème de la qualité des données et des services dans ses travaux autour de l'archivage et de la veille du web et l'agrégation de flux de données et médias sociaux. Enfin, l'extraction d'information synthétique et sémantiquement pertinente à partir de flots d'information sera aussi abordée.

5 Evaluation du département : analyse SWOT

On peut faire l'analyse SWOT suivante du département :

- **Points forts** : complémentarité des compétences, coopérations entre équipes, forte activité de publication et édition, fort réseau international et national, centre commun de recherche Thales-Communications / UPMC CLEAR, contribution au laboratoire des usages en technologies d'informations numériques LUTIN.
- **Points faibles** : visibilité internationale d'équipes, mais pas du département, hétérogénéité et dispersion des centres d'intérêt applicatifs.
- **Risques liés au contexte** : départ à prévoir de B. Bouchon-Meunier, départs potentiels des maîtres de conférences HDR.
- **Possibilités liées au contexte** : domaines applicatifs prioritaires sur le plan national : numérisation et préservation des données, big data, aide au maintien à domicile et e-santé, participation à deux labex (SMART, OBVIL), participation aux axes prospectifs du laboratoire sur la e-santé, le big data, et le smart home.



Vague D : Campagne d'évaluation 2012-2013

5.2 Projet 2014–2018 - Equipe BD

1 Membres permanents de l'équipe

Nom	Prénom	Statut	Etablissement
Doucet	Anne	PU	UPMC
Amann	Bernd	PU	UPMC
Constantin	Camelia	MC	UPMC
Gançarski	Stéphane	MC	UPMC
Le Pape	Cécile	MC	UPMC
Naacke	Hubert	MC	UPMC
Piowowski	Benjamin	CR CNRS	UPMC

2 Évolution de l'équipe

L'équipe est composée de 7 permanents dont 3 sont habilités à diriger des recherches. Un membre de l'équipe (Hubert Naacke) possède déjà le dossier nécessaire pour soutenir son HDR et d'autres membres pourraient être candidats dans les 5 ans à venir. On prévoit également que Bernd Amann prendra la responsabilité de l'équipe en 2014.

3 Caractérisation de la recherche

Les recherches de l'équipe portent sur des aspects fondamentaux en bases de données (qualité, cohérence, optimisation) en s'appuyant sur des problématiques d'actualité de gestion de données à large échelle, rencontrées notamment dans les applications du Web 2.0.

4 Projet scientifique

Le projet scientifique de l'équipe s'inscrit dans la continuité des travaux en cours et dans les thématiques de recherche prioritaires du département DAPA. L'équipe oriente ses recherches en s'ouvrant vers des domaines connexes et complémentaires (recherche d'information) et en considérant de nouveaux axes de recherche et de nouvelles applications (médias sociaux, préservation numérique, cloud communautaire). L'axe représentation des données textuelles correspond à l'arrivée de B. Piowowski dans l'équipe. Ces nouvelles orientations sont liées aux défis du département DAPA indexation à large échelle, distribution, dynamique et aux domaines d'application recherche d'information, médias sociaux et préservation de données.

L'équipe possède une grande expérience dans la représentation de données complexes (données sémantiques, documents Web/XML, flux de données) au niveau logique (sémantique, cohérence) et physique (stockage, indexation, distribution). Cette expérience permet d'étudier des applications complexes (archivage du web, Web2.0, médias sociaux) et de contribuer à améliorer à la fois la qualité et la performance de ces applications.

5 Thème : Flux de Données et Médias Sociaux

Description Ce thème de recherche cible les problèmes de passage à l'échelle et de dynamique rencontrés dans le traitement et la personnalisation de flux de données Web 2.0. Nous proposons d'étudier en particulier le traitement de requêtes top- k continues dans les réseaux sociaux. Comparée à une requête classique qui retourne une réponse complète sur une base de données statique, une requête top- k continue est évaluée sur un ou plusieurs flux de données et maintient une fenêtre des k items les plus pertinents par rapport à une fonction de score donnée. Notre objectif est d'étudier des requêtes avec des fonctions de scores sociaux qui agrègent "en temps réel" des événements d'utilisateurs dans un réseau (avis, clicks, ...). Ce type de requête est fondamental dans de nombreuses applications de type "filtrage collaboratif". La dynamique des informations impliquées (données, événements) et la complexité des fonctions de scores utilisées nécessitent des approches d'optimisation nouvelles qui combinent différentes techniques de stockage, d'indexation et de réplication.

Originalité L'originalité de notre projet se situe dans le rapprochement des fonctions de classement étudiées en recherche d'information (similarité, pertinence, importance) avec des techniques bases de données (optimisation de requêtes, transactions distribuées) pour le filtrage et la personnalisation efficace et expressive de flux d'informations générées dans les médias sociaux.

Positionnement Ces travaux envisagés s'inscrivent dans les axes de recherche prioritaire du département DAPA et dans le projet CNRS interdisciplinaire MASTODONS ARESOS (Reconstruction, Analyse et Accès aux Données dans les grands réseaux Socio-Sémantiques). Le filtrage et l'agrégation d'informations dans les réseaux sociaux posent des problèmes riches et complexes qui concernent plusieurs domaines informatiques (bases de données, apprentissage, recherche d'informations, modélisation de graphes) et qui est au centre des recherches de nombreux laboratoires au niveau national (LIG, INRIA, IRIT) et international (Max-Planck, EPFL, MIT). Dans ce contexte très compétitif, la particularité de notre approche se situe dans l'étude combinée des techniques de maintenance de scores dynamiques et d'optimisation de requêtes top- k appliquées aux flux de données.

Objectifs de résultats à 5 ans Nous avons déjà démarré une thèse (N. Vouzoukidou) avec des premiers résultats encourageants. Notre objectif à 5 ans est l'obtention d'une forte visibilité de notre activité à travers des publications et des contrats de recherche en renforçant les collaborations existantes (équipe MALIRE, CNAM, Univ. de Crête) en créant des liens nouveaux avec d'autres équipes et partenaires industriels.

6 Thème : Archivage du Web et Préservation Numérique

Description Les perspectives du thème à court terme sont bien évidemment liées aux thèses et au projet européen en cours: recherche d'information dans les archives et segmentation/comparaison hybride de pages Web (voir partie bilan). À moyen terme nous souhaitons explorer d'autres dimensions (que la temporelle) de la qualité des archives, notamment la dimension spatiale : comment s'assurer qu'une archive est un bon échantillon d'un domaine (géographique, thématique) ? Comment découvrir de nouvelles pages ou de nouveaux sites à inclure pour améliorer cette qualité ? Par ailleurs, le nombre grandissant d'archives (nationales, régionales, privées, thématiques), chacune forcément incomplète (temporellement et spatialement) incite à exploiter cette diversité en construisant des fédérations d'archives, dont il faut étudier le problème de leur constitution aussi bien que celui de leur exploitation.

Originalité Nous étudions conjointement la construction de l'archive et son exploitation par navigation et surtout recherche d'information, ce qui permet d'adapter la construction à l'utilisation ultérieure (très peu d'équipes dans le monde s'intéressent déjà à la RI dans les archives). D'autre part, nous basons nos travaux sur une perception fine de la structure et du contenu des pages (notamment en mêlant techniques d'analyse structurelle et celles d'analyse d'image), ce qui permet d'obtenir un comportement plus précis.

Positionnement Notre activité s'inscrit parfaitement dans les défis et applications du département DAPA, ainsi que dans l'axe transverse LSDP. Au niveau national, nous sommes pratiquement la seule équipe académique à nous intéresser aux problèmes d'archivage du Web, hormis la BNF et l'INA qui ont mission de dépôt légal, mais dont les travaux sont plus appliqués et ne donnent quasiment pas lieu à publication. Au niveau international, notre implication dans le projet européen Scape (projet européen IP de grande taille) nous garantit d'être intégré à la communauté internationale du domaine.

Objectifs de résultats à 5 ans Notre objectif est principalement de consolider notre approche afin d'envisager de construire une plate-forme complète d'archivage fédérée. Cette plate-forme pourra être déclinée, selon les besoins, en une version institutionnelle (archive de grande taille) et en une version individuelle. Dans les deux cas, la capacité de la plate-forme à faire collaborer plusieurs archives pose des problèmes à la fois d'infrastructure "système" et de sémantique afin de mettre en correspondance les différents contenus des archives.

7 Thème: Gestion de Données Complexes Distribuées

Description Les recherches du thème s'inscrivent dans les défis du département DAPA indexation à large échelle et dynamité, données distribuées et dans les domaines d'applications des média sociaux. Il vise des problèmes d'accès à des données complexes dans les clouds communautaires où les utilisateurs, les données et les ressources de traitement sont géo-localisés. L'objectif est de proposer des nouvelles stratégies de répartition dynamique des données et des traitements, tout en équilibrant la charge sur les points de l'infrastructure par rapport à leurs ressources. Nos investigations abordent en particulier la fragmentation dynamique des données en fonction des demandes d'accès avec la conception d'un répertoire des répliques et d'un algorithme d'évaluation répartie et partagée des requêtes en minimisant les ressources utilisées.

Les applications visées sont les services géo-localisés de réalité augmentée et l'accès à la base de données de biodiversité du GBIF.

Originalité La notion de service de gestion de données dans un cloud communautaire est récente. L'originalité de notre projet réside dans la prise en compte de la forte dispersion géographique des ressources qui nécessite des stratégies de regroupement dynamiques des données et des traitements. Nous avons une approche complémentaire des récentes solutions proposant des services de gestion de données sur le cloud (HadoopDB, Hadoop++, Relational Cloud, Albatross) qui font abstraction de la localisation des ressources.

Positionnement dans le contexte local/ national/ international Les travaux sur la distribution des données de bio-diversité s'effectuent dans le cadre d'une collaboration avec le consortium international du GBIF au Danemark et avec le MNHN qui est le correspondant GBIF pour la France. Nous avons également une collaboration avec l'UCAD à Dakar sur ce sujet, qui se traduit par une bourse de thèse du programme doctoral international UPMC PDI-MSc.

Objectifs de résultats à 5 ans Notre objectif est de concevoir les briques de base d'une plateforme communautaire pour la gestion distribuée de données géo-localisées.

8 Thème : Représentation des données textuelles

Description Il est nécessaire de développer des techniques de transformation automatique d'un texte vers un graphe. Afin d'atteindre cet objectif, nous nous appuyons sur trois domaines qui sont très actifs depuis ces dix dernières années: (1) la linguistique computationnelle, qui permet d'obtenir une représentation structurée d'une phrase; (2) le Web Sémantique qui permet un accès à un nombre toujours croissant de connaissances; (3) l'apprentissage structuré, qui permet d'envisager la simulation du processus de lecture comme une fonction itérative sur les graphes, dont les paramètres seraient à apprendre. Il sera ensuite nécessaire de développer des techniques pour ensuite exploiter ce type de représentation pour différentes tâches de l'Accès à l'Information (AI) (ex. recherche, résumé automatique).

Originalité Ce thème a pour but de développer une représentation des textes radicalement différente de ce qui est fait à l'heure actuelle dans le domaine de l'AI, où les documents sont représentés comme des graphes valués de quasi-prédicats qui ont été proposés en psycholinguistique comme étant de bons modèles cognitifs de représentation du contenu d'un texte. Ces modèles restent, à l'échelle de la RI, des modèles "jouet" car il n'est pas possible de traiter ("simuler sa lecture") automatiquement un texte, et encore moins de grandes quantités de textes.

Positionnement dans le contexte local/ national/ international Au niveau local, nos travaux s'inscrivent dans les domaines de recherche du Web sémantique et de l'apprentissage. Au niveau national et international,

du fait qu'il n'existe pas d'approches similaires dans le domaine de l'AI, nous nous appuyons sur les développements récents en linguistique computationnelle (Univ. d'Edinburgh) et psycholinguistique (Univ. Joseph Fourier).

Objectifs de résultats à 5 ans Notre objectif est tout d'abord de développer des prototypes, afin de mieux cerner les difficultés et problèmes à résoudre, ainsi que les collaborations nécessaires. Dans un second temps, et en fonction des résultats et nécessités, nous essaierons d'obtenir un financement pour une ou plusieurs thèses, post-doctorats, et/ou postes d'ingénieurs.

9 Analyse SWOT

Points forts :

- défis importants et d'actualité
- recherche portée par des applications réelles
- complémentarité des compétences dans l'équipe

Points faibles :

- objectifs ambitieux par rapport à la taille de l'équipe
- peu de contacts industriels dans certaines thématiques

Risques liés au contexte :

- domaine de recherche très évolutif
- contexte concurrentiel au niveau industriel
- faible adéquation des thématiques de recherche avec les appels à projets

Possibilités liées au contexte :

- collaboration avec d'autres équipes du département et du laboratoire
- attractivité de l'environnement pour les étudiants et les chercheurs
- proximité des structures de soutien à la recherche (pôles de compétitivité, LABEX, IDEX)

10 Mise en œuvre

Valorisation et recherche de moyens : L'équipe est actuellement engagée dans un contrat Européen (SCAPE) jusqu'en 2014 et dans une action CNRS (Mastodons ARESOS) acceptée en 2012 et prévue pour 1 an. On prévoit de répondre aux futurs appels à projets (ANR, pôle de compétitivité). Nous espérons également renforcer nos collaborations avec nos partenaires industriels actuels (France Télécom, Internet Memory, EADS, INA, Antidot).

Diffusion des résultats (stratégie de publication) : Nos projets de recherche à 5 ans s'inscrivent dans la continuité des travaux actuels. Plusieurs thèses et contrats de recherche sur ces thématiques sont déjà en cours et notre objectif est de consolider les premiers résultats avec des publications dans des conférences (CIKM, Data Engineering, EDBT) et des journaux de référence (DKE, TWEB, TOIT). Parallèlement, on envisage d'organiser et de participer à des ateliers plus spécialisés pour rendre nos travaux accessibles à d'autres communautés et pour créer des liens de collaborations nouveaux.

Stratégie locale / nationale/ internationale : On prévoit de jouer un rôle d'animation dans la communauté par la participation dans le GDR I3 et par l'organisation de conférences et d'écoles thématiques. Au niveau national nous sommes engagés dans l'action nationale CNRS Mastodons ARESOS et on prévoit la participation à d'autres actions dans le cadre des pôles de compétitivité régionaux et de futurs LABEX. Nous envisageons également de renforcer et d'étendre nos contacts au niveau Européen, mais également avec des pays comme le Venezuela et le Brésil. L'équipe participe au GDR International Brésil.

Implication future dans l'enseignement : L'équipe coordonne l'ensemble de l'enseignement de bases de données à l'UPMC, et participe à toutes les unités d'enseignement en BD en Licence et en Master Informatique. B. Amann sera co-responsable de la spécialité "Données, Apprentissage, Connaissances" dans la nouvelle habilitation du master. L'équipe est également impliquée dans le Master Erasmus Mundus DMKM.

11 Références

- [1] K. Munagala, U. Srivastava, and J. Widom. Optimization of continuous queries with shared expensive filters. *PODS*, page 215, 2007.
- [2] S. Ranu. Answering top-k queries over a mixture of attractive and repulsive dimensions. *Proceedings of the VLDB Endowment*, (1):169–180, 2012.
- [3] Z. Zhang, S.-w. Hwang, K. C.-c. Chang, M. Wang, C. A. Lang, and Y.-c. Chang. Boolean + Ranking: Querying a Database by K-Constrained Optimization. In *ACM SIGMOD international conference on Management of data*, pages 359–370, New York, New York, USA, June 2006. ACM Press.
- [4] Y. Zhou, A. Salehi, and K. Aberer. Scalable delivery of stream query result. *Proceedings of the VLDB Endowment*, 2(1):49–60, 2009.



Vague D : Campagne d'évaluation 2012-2013

5.3 Projet 2014–2018 - Equipe LFI

1 Membres permanent de l'équipe

Nom	Prénom	Statut	Etablissement
Bouchon-Meunier	Bernadette	DR	CNRS
Detyniecki	Marcin	CR HDR	CNRS
Gacogne	Louis	PRAG	ENSIIE
Labroche	Nicolas	MC	UPMC
Lesot	Marie-Jeanne	MC	UPMC
Marsala	Christophe	MC HDR	UPMC
Rifqi	Maria	MC HDR	Univ. Paris 2
Tollari	Sabrina	MC	UPMC

2 Evolution de l'équipe

L'équipe LFI (Learning, Fuzzy and Intelligent systems) est issue de la composante "apprentissage dans l'incertain" de l'équipe MALIRE.

Dans cette équipe, le départ en retraite de B. Bouchon-Meunier est à prévoir dans l'année universitaire 2013–2014. Un poste de professeur des universités est demandé par l'équipe pour un recrutement en 2013.

3 Caractérisation de la recherche

Nos recherches ont pour cadre l'utilisation de techniques d'intelligence computationnelle, telles l'apprentissage artificiel ou la logique floue, pour la conception de systèmes intelligents. Pour cela, nous réalisons à la fois des recherches fondamentales, pour la conception de nouveaux modèles ou l'extension de modèles théoriques existants pour la prise en compte de connaissances graduelles ou imparfaites, mais aussi des recherches dans des domaines applicatifs tels le multimédia, l'internet des objets, la diffusion d'information textuelle sur le web en particulier.

4 Projet scientifique

L'originalité de l'équipe LFI réside dans ses recherches théoriques sur des méthodes d'apprentissage inductif et de clustering dans l'incertain alliées à sa capacité à traiter des données subjectives ou sensorielles, ainsi que sur des méthodes de modélisation du raisonnement et de la connaissance basées, entre autres, sur la logique floue. Les applications visées ont une forte implication sociétale et sont au cœur de la vie numérique actuelle. Parmi les équipes françaises travaillant sur les mêmes problématiques fondamentales, LFI présente la particularité de couvrir un large champ d'applications, en particulier dans des collaborations avec des partenaires universitaires ou industriels. L'équipe est reconnue au niveau international et son projet s'insère dans les axes de recherche développés en particulier en Allemagne, en Pologne, en Espagne, en Italie, au Japon ou aux Etats-Unis.

De nouvelles thématiques de recherche seront développées, telles que la prise en compte de la temporalité, de la taille des flots de données et des informations subjectives inhérentes au domaine de l'affective computing, ainsi que par divers domaines abordés, liés à l'évolution des pratiques numériques et à l'importance prise par des sujets cruciaux, tels que la e-santé ou le maintien à domicile.

La prise de risque se situe essentiellement à deux niveaux : d'une part la mise au point d'algorithmes nouveaux prenant en compte les aspects temporels et la taille considérable des données ; d'autre part la reconnaissance et le traitement d'informations sensorielles ou émotionnelles nécessitant une liaison difficile entre données objectives issues de tous types de capteurs et données subjectives ou ressenties. Les axes de prospective transverse du LIP6 tels que la e-santé, le large scale et l'intelligence ambiante sont étroitement liés à la prospective de l'équipe LFI, ainsi que l'axe "l'Homme à la convergence des environnements réel et numérique" que nous coordonnons dans le labex SMART.

4.1 Apprentissage supervisé et non supervisé dans l'incertain

Description. Les méthodes d'apprentissage abordées par l'équipe LFI concernent l'apprentissage supervisé (en particulier par arbres de décision flous) et l'apprentissage non supervisé (par des méthodes de clustering floues). Elles seront abordées sous l'angle de la temporalité pour traiter tout système évolutif dans un environnement dynamique, par exemple pour pouvoir traiter des flux d'information web, des réseaux de sources évoluant continuellement, ou encore des objets internet actifs. En particulier, des méthodes d'apprentissage incrémental seront développées ; cette incrémentalité, en apprentissage supervisé comme non supervisé, constitue en outre un facteur clef de passage à l'échelle. De plus, hors l'aspect temporel, d'autres éléments de contextualisation seront également pris en compte, tels que la localisation ou l'environnement social. Des caractéristiques plus spécifiques de certains problèmes, comme la monotonie des classes relativement à des valeurs d'attributs ou la mise en évidence de gradualité entre des variables seront aussi explorées. Temporalité et contextualisation sont deux composantes incontournables de bien des systèmes évolutifs et participent à des recherches très actuelles.

Originalité. L'originalité des travaux de LFI réside dans leur utilisation de la théorie des sous-ensembles flous et de la logique floue. Le traitement de gradualité a déjà été abordé dans plusieurs directions par des membres de LFI et constitue un axe de recherche prometteur.

Positionnement dans le contexte local / national / international. La temporalité et le passage à l'échelle sont deux des axes sur lesquels travaillent les chercheurs en apprentissage automatique, en particulier dans les autres équipes du département DAPA, ainsi que dans le département RSR. Des travaux dans les directions évoquées existent dans de nombreux laboratoires nationaux (LIRMM, Télécom ParisTech, INRIA,...) ou internationaux (universités de Magdeburg, du Missouri, Carnegie Mellon,...) avec qui nous collaborons déjà, comme le montrent nos publications communes.

Objectifs de résultats à 5 ans. L'objectif de l'équipe LFI sur les 5 années à venir est d'une part de développer de nouvelles approches d'apprentissage incrémental prenant en compte la temporalité des données et leur contexte, et d'autre part, de les mettre en œuvre dans des domaines applicatifs (multimédia, internet des objets...) afin d'en vérifier le passage à l'échelle et l'efficacité.

4.2 Modélisation du raisonnement et de la connaissance

Description. Ce sujet est dans la continuité des travaux antérieurs des membres de l'équipe. Les avancées attendues dans les années à venir concernent les aspects du raisonnement approximatif liés aux problématiques mentionnées : temporalité, complexité des informations et passage à l'échelle. Elles se focaliseront sur des méthodes de fusion de données hétérogènes et imparfaitement connues, sur des méthodes de raisonnement interpolatives ou graduelles, permettant de traiter des masses de données sur lesquelles existent peu de connaissances. Elles concerneront aussi la construction de résumés linguistiques, afin de fournir une information synthétique et aisément compréhensible à partir de grandes masses de données qui restent difficilement appréhendables par l'utilisateur.

Originalité. L'originalité de la construction de résumés linguistiques réside dans l'utilisation de modélisation floue alliée à de l'apprentissage et elle sera renforcée par une coopération avec des linguistes pour plus de richesse d'expression des résultats.

Positionnement dans le contexte local / national / international. Les travaux sur le raisonnement approximatif s'insèrent naturellement dans les axes français et internationaux sur la logique floue, comme en témoignent

les conférences plénières des divers membres de LFI ces dernières années et leur implication dans l'organisation de conférences nationales (LFA) ou internationales (SSCI, Fuzz-IEEE,...). Des contacts ont été pris avec le laboratoire ALPAGE (Université Paris Diderot) sur les résumés linguistiques. Ceux-ci constituent un axe de recherche actuellement très actif sur le plan international, illustré par exemple par un projet européen en cours de montage auquel nous sommes associés avec, comme coordinateur, l'European Centre for Soft Computing (Espagne) et des partenaires européens tels que la Polish Academy of Science (Pologne), avec qui nous collaborons déjà.

Objectifs de résultats à 5 ans. Sur ce thème, l'objectif à 5 ans est de développer de nouvelles approches formelles, ou des extensions d'approches existantes, permettant la prise en compte de connaissances graduelles ou imparfaites. Ces approches seront appliquées dans nos domaines applicatifs afin de mieux en étudier les propriétés et de pouvoir mettre en œuvre leur validation expérimentale.

4.3 Applications sociétales

Description. Les domaines d'application en perspective concernent des problèmes dans lesquels informations numériques et composantes humaines sont étroitement liées, ce qui rend leur traitement difficile, mais vital pour l'environnement à prévoir dans les années à venir. Parmi ces domaines, on peut citer la e-santé sous les aspects de traitement de données physiologiques en liaison avec l'état mental ou émotionnel des patients, ou encore le smart home en ce qui concerne essentiellement le maintien à domicile des personnes âgées ou en situation de handicap. Sont également concernés des domaines liés à la diffusion massive des informations sur internet, et en particulier son utilisation pour l'analyse de traces utilisateurs, mais aussi pour gérer des dispositifs physiques dans le domaine de l'internet des objets, ou l'étude de la dynamique informationnelle et des problèmes de suivi, d'évolution et de sécurité des informations qui lui sont liés. La détection de rumeurs, la cotation des informations comme évaluation de leur qualité ou de leur fiabilité, constituent des axes importants pour la vie économique, culturelle, sociale ou politique, qui seront abordés. D'autres domaines d'application seront également explorés, tels que la conception de vêtements intelligents ou la recherche intelligente d'images ou plus généralement d'informations multimédias, en particulier en utilisant des critères subjectifs ou émotionnels.

Originalité. L'originalité de l'approche de ces problèmes par l'équipe LFI tient, ici encore, à l'association entre apprentissage automatique et traitement de l'incertain, souvent par des méthodes floues ou possibilistes dans le cadre de l'intelligence computationnelle. Ces méthodes s'avèrent particulièrement intéressantes pour prendre en compte des composantes humaines tout en traitant des masses de données factuelles ou numériques.

Positionnement dans le contexte local / national / international. Le domaine de l'intelligence computationnelle (affective computing) fait l'objet, en particulier, d'une collaboration avec l'équipe SYEL, celui de la dynamique informationnelle avec des membres de l'équipe MLIA. Sur le plan national, ces recherches rejoignent d'autres poursuivies à l'Institut Télécom ParisTech sur l'affective computing, ceux développés à l'ENSAM avec qui nous coopérons à travers des projets communs, ainsi que les travaux du laboratoire des usages LUTIN avec qui nous avons, en particulier, récemment participé à un dépôt de projet européen sur le Smart Ambient Living Lab et avec qui nous collaborons dans divers domaines liés à l'utilisation de capteurs sensoriels. La cotation des informations et l'étude de la dynamique informationnelle sont au cœur de nos relations avec Thales Communications, actuellement au sein de l'équipe commune CLEAR.

Objectifs de résultats à 5 ans. Pour ce thème essentiellement applicatif, notre objectif à 5 ans est de pouvoir utiliser nos approches formelles dans un cadre réel d'utilisation. Une telle mise en œuvre nous permettra à la fois de vérifier la validité de nos approches, mais aussi d'alimenter notre recherche fondamentale par les problèmes soulevés lors de l'expérimentation *in situ* de nos méthodes.

5 Analyse SWOT

Points forts : L'équipe LFI exploite ses doubles compétences d'apprentissage et de traitement d'incertitude, en particulier par la logique floue dont elle est une des équipes majeures au niveau national et qui fait partie des équipes internationalement reconnues. Les modélisations floues sont particulièrement efficaces dans tous les problèmes imprécis, traitant des facteurs humains et manipulant des données subjectives et sont donc particulièrement adaptées aux domaines mis en perspective.

Points faibles : Le point faible principal de l'équipe LFI se situe au niveau de la gestion des ressources humaines liée au départ à la retraite de B. Bouchon-Meunier en 2013-2014.

Risques liés au contexte : Sur un plan technique, la grande variabilité des indicateurs biologiques ou sensoriels dans une population donnée rend difficile une approche générale de l'identification de facteurs émotionnels ou psychologiques. Sur un plan applicatif, les domaines sociétaux mentionnés sont largement explorés, de façon multidisciplinaire.

Possibilités liées au contexte : En France, l'équipe peut compter sur ses collaborations pluridisciplinaires, en particulier avec les psycho-cogniticiens du laboratoire des usages LUTIN et avec des linguistes, ainsi qu'avec des industriels comme Thales Communications. Sa participation au Labex SMART sera indubitablement la source de collaborations enrichissantes sur le plan local. Elle participe également à des groupes de travail internationaux en particulier au sein de la IEEE Computational Intelligence Society, et lie d'étroites collaborations scientifiques avec l'European Center for Soft Computing.

6 Mise en œuvre

Valorisation et recherche de moyens : Dans l'état actuel, l'équipe participe à des projets avec l'équipe CLEAR (Thales Communication & LIP6) avec laquelle elle réalise aussi des réponses à des appels à projets. Dans les années à venir, l'équipe s'appuiera sur son réseau de collaborations nationales et internationales afin de participer à des projets en Europe, avec le Japon, le Vietnam ou le Brésil (certains ont déjà été déposés ou sont en préparation).

Diffusion des résultats (stratégie de publication) : La publication dans des revues internationales sera renforcée. L'équipe s'attachera à diffuser ses travaux dans des communautés scientifiques diverses, conformément à la pluralité de ses recherches. D'autre part, l'équipe proposera des tutoriaux dans des conférences nationales ou internationales.

Stratégie locale / nationale/ Internationale : L'équipe renforcera ses relations avec des équipes multidisciplinaires en France : linguistes d'ALPAGE (université Paris Diderot), équipes médicales de l'UPMC, psycho-cogniticiens de l'université Paris 8, spécialistes de Big Data de l'université Montpellier II, ainsi qu'avec des entreprises dont les compétences sont complémentaires des siennes, en particulier grâce à son implication dans le pôle de compétitivité Cap Digital. L'équipe renforcera son leadership sur ses sujets d'excellence en continuant à organiser la conférence internationale IPMU et en animant des workshops internationaux dans la continuité de ceux qu'elle a créés ces dernières années (Affective Computational Intelligence, Adaptive Multimedia Retrieval, Medical and Educational Applications of Computational Intelligence to Benefit Society).

Implication future dans l'enseignement : Les membres de l'équipe participent activement à l'enseignement (UPMC, Paris 2, ENSIEE, Paris 8). En particulier, nous sommes impliqués dans le montage des enseignements du L et du M pour la prochaine habilitation avec la proposition et la responsabilité de plusieurs UE tant au niveau L (UEs d'initiation à l'informatique, de Logique, ou d'intelligence artificielle) qu'au niveau M (UEs d'intelligence artificielle, ou de modélisation du raisonnement). L'équipe intervient aussi dans le Master européen DMKM.



Vague D : Campagne d'évaluation 2012-2013

5.4 Projet 2014–2018 - Equipe MLIA

Limiter le total (hors références) à 3 pages - Au plus 5 références pour le positionnement

1 Membres permanent de l'équipe

Prénom	Nom	Statut	Etablissement
Thierry	Artières	PU	UPMC
Nicolas	Baskiotis	MC	UPMC
Matthieu	Cord	PU	UPMC
Ludovic	Denoyer	MC	UPMC
Patrick	Gallinari	PU	UPMC
Vincent	Guigue	MC	UPMC
Sylvain	Lamprier	MC	UPMC
Nicolas	Thome	MC	UPMC
Nicolas	Usunier	MC	UPMC

2 Evolution de l'équipe

L'équipe MLIA ("Machine Learning and Information Access") constitue la composante apprentissage statistique de l'équipe MALIRE que nous avons choisi de séparer en deux pour une meilleure visibilité. Pour cette nouvelle équipe, plusieurs départs sont à prévoir sur les 5 prochaines années. M. Amini MC, est nommé professeur à Grenoble au 1er octobre 2012, plusieurs MC passeront leur habilitation sur la période à venir. Assurer la continuité du dynamisme de l'équipe passera par de nouveaux recrutements.

3 Caractérisation de la recherche

L'apprentissage statistique constitue le domaine fondamental et le coeur du travail de l'équipe. Notre positionnement est algorithmique tout en nous appuyant sur des aspects fondamentaux et applicatifs. Nous défendons une vision applicative du domaine et pour cela nous investissons sur des champs d'application privilégiés qui sont la recherche d'information texte et multimédia, la fouille de réseaux sociaux.

4 Projet scientifique

Le projet s'articule sur 4 thèmes. Le développement de techniques d'apprentissage pour le traitement de contenus sémantiques constitue l'axe applicatif majeur. Deux thèmes algorithmiques portent l'un sur le traitement de données relationnelles, l'autre sur apprentissage et optimisation. Un dernier thème sera centré sur le problème spécifique de la conception d'algorithmes pour des données massives et très large échelle et comporte des aspects algorithmiques et applicatifs.

- Evolutions par rapport à la période précédente /Prise de risque

Les trois premiers thèmes correspondent à des recherches sur lesquelles l'équipe est déjà positionnée ou qu'elle a commencé à explorer. Le dernier thème focalisé sur le très large échelle est nouveau.

- Positionnement par rapport aux thèmes transverses et de département

Le thème large échelle entre entièrement dans le thème transversal “large Scale Data Processing” du laboratoire. L’équipe sera également fortement présente sur le thème transversal “Social Computing” du laboratoire par ses travaux sur les données relationnelles. Les 4 thèmes entrent dans les directions de recherche du Labex SMART.

4.1 Traitement de contenus

Description Sur le traitement de contenus, l’équipe se positionne sur la conception d’algorithmes d’apprentissage pour des tâches de recherche d’information (RI) émergentes. Une première direction concerne les modèles d’apprentissage pour la RI. Après avoir participé activement aux travaux sur la RI XML, nous débutons un projet de recherche sur la RI sur des flux sociaux. Les challenges pour l’apprentissage sont le traitement de données bruitées, fortement dynamiques et relationnelles, et le développement d’outils pour les nouveaux besoins qui ont émergé en RI (crowd sourcing, prise en compte de multiples sources d’information, analyse de sentiments, recommandation et diffusion d’information, RI sociale). Une seconde direction concerne la représentation des données textuelles et visuelles : comment représenter ces données complexes et en très grande dimension pour un ensemble de tâches de RI. Nous étudierons des approches basées sur l’apprentissage d’architectures originales de réseaux de neurones : les réseaux profonds, et les réseaux à convolutions. L’approche est à la fois en rupture et complémentaire des méthodes récemment développées comme les méthodes à noyaux et celles à variables latentes. Elle offre un potentiel de représentation riche mais pose encore de nombreux problèmes d’estimation et d’apprentissage. En vision, s’ajoute une source d’inspiration liée aux systèmes bio-inspirés qui pourra motiver également certaines de nos architectures. Nous développons enfin également une collaboration avec l’équipe BD sur des problèmes de représentation de texte par des graphes en utilisant de l’apprentissage structuré et des sources d’information multiples.

Originalité L’équipe a été pionnière sur le thème “apprentissage et RI” en lançant cette direction de recherche il y a une dizaine d’années tout en se positionnant sur des domaines émergents de la RI. Aujourd’hui alors que l’apprentissage est devenu omniprésent en RI, nous avons par notre expertise une position privilégiée pour contribuer à ce domaine par de nouvelles approches algorithmiques et en adaptant pour la RI des méthodes très récentes de l’apprentissage. L’analyse des flux sociaux est un domaine émergent, même si de nombreuses équipes essaient de s’y positionner. L’analyse par des représentations neuronales est encore assez peu développée et l’équipe a des compétences sur le domaine.

Positionnement dans le contexte local/ national/ international En France il existe des communautés importantes RI et multimédia dans lesquelles nous sommes bien insérés. Sur le positionnement apprentissage et texte, l’équipe d’E. Gaussier au LIG a un positionnement similaire au notre, celle de F. Yvon au LIMSI a des compétences proches mais s’est orientée sur la traduction. Des équipes majeures sur ces domaines existent dans de grandes universités avec des groupes issus de l’apprentissage (A. McCallum à CMU), de la vision (M. Hebert CMU) ou même des bases de données (G. Weikum Max-Planck). Il existe peu d’équipes Françaises travaillant sur les réseaux de neurones profonds. Le LIP6 est impliqué sur un projet ANR ASAP dirigé par A. Rakotomamonjy (LITIS, Rouen) sur cette problématique. Parmi les équipes phares du domaine, on trouve celles de G. Hinton et Y. Bengio au Canada, ainsi que des équipes américaines (Y. LeCun, NYU ; A. Ng, Stanford ; T. Poggio, MIT sur les aspects bio-inspirés).

Objectifs de résultats à 5 ans L’objectif est de renforcer notre positionnement international sur l’analyse de flux sociaux et sur les outils spécifiques comme les réseaux de neurones profonds pour le texte et l’image, et de développer des collaborations internationales sur ces deux aspects.

4.2 Apprentissage relationnel

Description Le domaine de l’apprentissage s’étant surtout focalisé sur le traitement de données vectorielles, l’émergence de sources de données relationnelles (documents structurés, Web, réseaux sociaux) nécessite de nouvelles méthodes. Nous positionnons nos travaux sur le traitement simultané de relations complexes entre les données et de contenu associé à ces données.

La complexité provient de la nature dynamique ou multi-type des relations et de l’hétérogénéité des données. Cette prise en compte simultanée du contenu (texte, image, ... voir thème précédent) et de la

structure relationnelle est relativement peu étudiée bien que primordiale pour le traitement d'applications de type Web ou réseaux sociaux par exemple. Nous planifions d'étudier ces aspects à travers trois pistes de recherche principales :

l'apprentissage de prédiction structurée sous faible supervision en revisitant différentes instances de cette problématique (apprentissage semi-supervisé, one shot learning, transfer learning). Nous focaliserons nos travaux principalement sur les données séquentielles qui sont très répandues (sons, parole, musique, signaux oculométriques, capteurs de mouvements, signaux biomédicaux, etc) et pour lesquelles, La diffusion de grands volumes de données et la multiplication des besoins applicatifs rendent irréaliste la création de grands corpus étiquetés.

Les méthodes d'apprentissage séquentiel qui consistent à considérer le processus d'inférence - classification par exemple - comme un processus séquentiel qui, à chaque étape, décide d'explorer une partie des données afin de décider. Ce sont des méthodes prometteuses, tant du point de vue de la vitesse d'inférence (voir thème 4), de leur capacité à traiter tout type de données relationnelles, mais aussi, dans leur possibilité à résoudre des tâches complexes comme la transformation de structures par exemple.

Le développement de méthodes pour le traitement de données relationnelles dynamiques. Particulièrement, nous étudierons des problèmes de prédiction de propagation d'information dans les réseaux, avec en ligne de mire, des applications autour du traitement de réseaux sociaux, de la recommandation sociale et de réseaux de capteurs.

Originalité Les principales originalités dans les thèmes abordés sont : la prise en compte d'une supervision très faible pour le traitement de données structurées qui n'a suscité que peu de travaux, l'utilisation de modèles séquentiels d'apprentissage et leur application pour des applications récentes comme la propagation d'informations dans les réseaux sociaux. Si ces trois axes de recherche constituent chacun un problème formel intéressant, nous considérons, en lien avec la perspective sur le traitement de grandes masses de données, que conjointement, ces directions de recherche permettront le développement de modèles efficaces pour résoudre des problèmes réels impliquant de grandes masses de données.

Positionnement dans le contexte local/ national/ international Les paradigmes d'apprentissage sous faible supervision ont suscité beaucoup d'intérêt pour le traitement de données vectorielles et restent actuels (e.g. workshops à ICML 2011, NIPS 2001), mais les travaux sur des données structurées restent limités. Quelques travaux théoriques ont exploré le domaine comme (Brefeld et al., ICML 2006) ou (Chang et al., ICML 2010) mais sans application à des problèmes réels de grande taille.

L'utilisation de méthodes séquentielles pour la résolution de problèmes classiques d'apprentissage est une approche de plus en plus étudiée dans le contexte international, même si le nombre de travaux reste aujourd'hui assez faible. Citons les travaux de Balazs Kegl à l'université d'Orsay pour le traitement d'image, avec qui nous sommes en relation, ou bien encore les travaux de Hal Daume III à l'université du Maryland ou ceux de T. Rückstie au laboratoire "biomimetic robotics and machine learning. Les travaux effectués au LIP6 sont précurseurs dans le domaine.

Dans le cadre de la propagation de contenu dans les données relationnelles, nous travaillons aujourd'hui conjointement avec des industriels (Bell-Labs, BlogSpirit) et des universitaires (E. Gaussier, LIG).

Objectifs de résultats à 5 ans Si nous envisageons d'obtenir plusieurs résultats formels dans les années à venir, nous souhaitons particulièrement promouvoir des résultats applicatifs concrets. Ainsi, par exemple, nous sommes en train d'appliquer les méthodes séquentielles à des problématiques de vision (classification d'image, détection d'objet). Une application à moyen terme serait la résolution de problématique d'accès à l'information sur le Web et les réseaux sociaux par des méthodes séquentielles imitant le comportement humain : parcours du Web, interaction avec des utilisateurs réels.

4.3 Apprentissage et optimisation

Description Notre projet consiste à étudier les critères d'apprentissage et les algorithmes d'optimisation correspondants, pour les nouveaux défis posés par les données actuelles (dynamisme, large échelle, hétérogénéité) et pour un ensemble de besoins émergents. Une première direction concerne l'apprentissage d'ordonnancement d'exemples pour lequel nous avons déjà quelques résultats fondamentaux. Cette problématique est devenue cruciale pour des applications de l'apprentissage aux moteurs de recherche et

aux systèmes de recommandation. Nous allons d'une part développer des analyses théoriques et expérimentales sur les coûts de substitution convexes, en se focalisant sur les problématiques liées à la prise en compte de la diversité de la liste ordonnée de résultats, ainsi que celles liées aux données manquantes. Nous souhaitons par ailleurs étudier les algorithmes d'optimisation en ligne, tout d'abord dans les cadres sur lesquels l'équipe possède déjà une expertise : le cadre des problèmes d'optimisation convexe, ainsi que le cadre des méthodes séquentielles et de l'apprentissage par renforcement.

Originalité L'originalité de notre projet réside dans les problématiques de définition de fonctions de coûts adéquates aux problèmes étudiés et dans la conception des algorithmes d'optimisation associés. Sur les coûts de substitution convexes, notre originalité est d'étudier les problématiques d'ordonnement, la plupart des travaux considèrent des tâches de classification.

Positionnement dans le contexte local/ national/ international Les coûts de substitution convexes sont surtout étudiés à l'international, en particulier dans les équipes de B. Williamson en Australie, de E. Hüllermeier en Allemagne et les travaux de J. Duchi à Berkeley. Les problématiques d'optimisation en ligne sont très largement étudiées dans la communauté. On pourra en particulier citer les travaux de K. Crammer et S. Shalev-Shwartz en Israël.

Objectifs de résultats à 5 ans D'ici 5 ans, nous devons avoir obtenu de nouveaux algorithmes validés théoriquement et expérimentalement sur des grandes masses de données dynamiques issues de moteurs de recherche ou de recommandation pour la problématique d'ordonnement.

4.4 Apprentissage large échelle

Description Le large échelle correspond au traitement de très grandes quantités de données et nécessite de réviser les algorithmes traditionnels de l'apprentissage qui sont conçus principalement pour un traitement en mémoire centrale. Suivant les applications, ces données peuvent venir sous la forme de flux, être disponibles dans des bases de données sous une forme distribuées, être fortement hétérogènes ou encore combiner toutes ces caractéristiques. Une première direction de recherche concerne la classification en très grand nombre de classe. Les classifieurs actuels ont été développés pour traiter des problèmes à quelques dizaines de classes alors qu'actuellement se posent des problèmes avec plusieurs milliers ou dizaines de milliers de classes (les bases de brevets (<http://www.wipo.int/classifications/ipc/en/>) contiennent environ 70 000 catégories, DMOZ un des gros repository du web, contient plus de 600 000 catégories, Wikipedia a plus de 20000 catégories reliées par différents types de relations). Nous avons engagé une collaboration sur ce thème avec plusieurs équipes qui se concrétise par un projet européen (FP7) et l'organisation d'un challenge international coordonné par l'institut NCSR Demokritos d'Athènes. L'objectif est de développer de nouvelles approches et de procéder à des évaluations large échelle. Une seconde direction est l'apprentissage sous contrainte de ressources limitées (budget learning), que nous prévoyons de développer à partir de 2014. Il s'agit de tenir compte dans la conception même des algorithmes des contraintes de ressource (calcul, accès aux données) qui ne sont pas prises en compte actuellement. L'apprentissage de données sous forme de flux hétérogènes constitue une troisième direction.

Originalité Les problèmes de très large échelle commencent à être abordés en apprentissage, mais ils s'agit d'un thème émergent. Les problématiques de très grand nombre de classe ont par exemple été abordés par Yahoo ! research et par quelques équipes que nous connaissons à travers des challenges et workshops que nous organisons en 2011 et 2012, et par l'organisation en 2011 et 2012. L'apprentissage sous budget a fait l'objet de quelques travaux mais ne s'est pas concrétisé comme une direction majeure jusqu'à ces toutes dernières années. Il en est de même pour l'apprentissage décentralisé.

Positionnement dans le contexte local/ national/ international Il s'agit pour nous globalement d'une nouvelle thématique de recherche. Nous avons des collaborations sur les problèmes de grand nombre de classes. L'apprentissage sous contrainte de budget a fait l'objet de quelques travaux notamment à CMU (A. McCallum), NYU (J. Attenberg), Alberta (R. Greiner) etc. Les travaux les plus proches de notre positionnement sont ceux proposés par P.D. Turney, au Canada. L'apprentissage distribué sur des flux ou bases hétérogènes a également été très peu étudié.

Objectifs de résultats à 5 ans Sur ces thématiques neuves, l'objectif est de développer à la fois de nouvelles compétences internes en s'appuyant sur des thèmes transversaux du laboratoire (cas de l'apprentissage distribué) et de se positionner au niveau international sur les autres.

5 Analyse SWOT

Points forts : L'équipe possède des compétences complémentaires sur les différentes thématiques qui seront abordées (apprentissage et RI par exemple). Nous nous appuyons sur des applications actuelles et concrètes du traitement de données au travers de relations industrielles. L'équipe a une position de pionnier sur certaines des thématiques abordées.

Points faibles : Au niveau des ressources humaines le maintien des compétences suite à des départs prévisibles est un enjeu crucial. Il y a une prise de risque importante liée à la nouveauté de certains sujets en particulier sur le large échelle.

Risques liés au contexte : Il existe une forte compétitivité internationale sur ces thématiques liées à des problèmes Big Data. Certaines directions ne peuvent se développer que par des collaborations académiques internationales et par des collaborations industrielles locales.

Possibilités liées au contexte : Nous avons développé des relations privilégiées avec des acteurs du Big Data comme Thales, et des PME. Plusieurs des thèmes abordés entrent bien dans la prospective transversale du laboratoire et dans le labex SMART.

6 Mise en œuvre

Valorisation et recherche de moyens : L'équipe est engagée dans un projet européen qui débute sur l'extraction d'information et sur les problématiques de très grands nombre de classes. Elle est également partenaire dans des projets nationaux sur certaines de ces thématiques prospectives. Elle collabore avec Thales Com dans le cadre de l'équipe CLEAR et participe à des dépôts de projets.

Diffusion des résultats (stratégie de publication) : La stratégie de l'équipe est de viser des conférences de niveau A et des workshops spécialisés. Nous continuerons à nous impliquer dans l'organisation de workshops et de challenges qui sont des lieux de rencontre et de diffusion.

Stratégie locale / nationale/ Internationale : Les thématiques de l'équipe sont au centre des directions de recherche du labex SMART. Nous coordonnons un projet qui a été retenu dans le cadre de l'appel CNRS pluri-disciplinaire MASTODONS sur les masses de données que nous poursuivrons si l'opportunité se présente. Nous poursuivrons également l'effort de collaborations internationales par le montage de projets et les échanges de chercheurs.

Implication future dans l'enseignement : L'équipe est impliquée dans le montage du projet L et M de l'UPMC avec la prise en charge de plusieurs UE, la co-responsabilité d'une spécialité de master, la participation au master international DMKM (Data Mining Knowledge Management).



agence d'évaluation de la recherche
et de l'enseignement supérieur

Unités de recherche

Vague D : Campagne d'évaluation 2012-2013

5.5 Projet 2014–2018 - Equipe ACASA

1 Membres permanent de l'équipe

Nom	Prénom	Status	Etablissement
Ganascia	Jean-Gabriel	PU	UPMC
Roux	Magali	DR	CNRS
Bourgnes	Gauvain	MC	UPMC
Jouis	Christophe	MC	Univ. Paris 3

2 Évolution de l'équipe

Avec le recrutement de Gauvain Bourgnès, la participation à deux projets *investissements d'avenir* sur le livre électronique et l'implication dans le Labex OBVIL, l'équipe ACASA déploie de plus en plus ses activités sur la thématique des *humanités numériques*.

3 Caractérisation de la recherche

L'équipe ACASA travaille sur la *modélisation cognitive* à l'aide de techniques d'*intelligence artificielle symbolique*, en particulier avec des techniques d'*apprentissage symbolique*, de *modélisation du raisonnement*, en particulier de modélisation de l'*abduction*, de la *déduction* et de l'*induction*, de *traitement sémantique* de l'information, de *représentation des connaissances* et de *fusion symbolique de données*. A cela s'ajoute un axe portant sur la *philosophie computationnelle*, ce qui inclut l'*épistémologie computationnelle*, en ayant recours à la modélisation de l'abduction, et l'*éthique computationnelle*.

Ces recherches fondamentales débouchent sur un certain nombre d'applications, en particulier :

- application à la santé en collaboration avec des hôpitaux, plus précisément interprétation des signaux de polysomnographie à l'aide de techniques de fusion symbolique d'information ;
- protection de l'intimité de la vie privée et recommandation avec des techniques de traitement sémantique de l'information ;
- éditions électroniques, avec des éditeurs, en faisant usage de techniques de traitement sémantique de l'information, de traitement de la langue naturelle, de fouille de texte et de cartographie de contenus.

4 Projet scientifique

- Originalité et positionnement nat/ Int. : les relations établies entre l'équipe ACASA, spécialisée en intelligence artificielle symbolique, et des équipes de littérature, de philosophie et d'épistémologie constitue une originalité tant au plan national qu'international. Le champ des humanités numériques est encore peu développé en France, en particulier dans le domaine de la littérature. Là encore, c'est une originalité.
- Évolutions par rapport à la période précédente /Prise de risque : l'équipe s'ancre fortement sur le thème des *humanités numériques* sur lequel elle s'était déjà progressivement engagée avec ses travaux en philologie computationnelle réalisés en collaboration avec l'ITEM et ses travaux en épistémologie computationnelle conduits avec l'équipe d'épistémologie de l'ENS. Par ailleurs, elle renforce le traitement sé-

mantique de l'information avec des travaux sur l'abduction qu'elle approfondit avec la venue de Gauvain Bourgne et des travaux sur l'annotation sémantique qui font, entre autre, appel aux techniques développées par Jean-Pierre Desclés, dont l'équipe doit assurer le suivi.

- Positionnement par rapport aux thèmes transverses et de département : l'équipe ACASA est impliquée dans deux thèmes transverses du laboratoire, l'un sur la *santé* avec un projet de polysomnographie sans fil réalisé en collaboration avec l'équipe SYEL, l'autre sur le "*social computing*", au sein du département DAPA, avec en particulier les questions de recommandation et de protection de la vie privée. A cela on doit ajouter le traitement de gros corpus de textes qui s'inscrit dans la thématique "Big Data" du département DAPA.

4.1 Thème 1 : humanités numériques

Description : cela recouvre l'évolution des humanités traditionnelles, c'est-à-dire des disciplines d'érudition et des "sciences de la culture", grâce à l'utilisation des technologies de l'information. En l'occurrence, nous abordons la critique littéraire, l'édition électronique et la philosophie computationnelle avec des outils d'intelligence artificielle.

Originalité : à la différence de beaucoup d'équipes de notre pays qui abordent les sciences cognitives comme des "sciences de la nature" visant à naturaliser les problématiques philosophiques à l'aide des ressources offertes par les technologies de l'information, l'équipe ACASA envisage les sciences cognitives sous l'angle des "sciences de la culture", dans le but d'étudier, de représenter et de simuler les raisonnements humains, en particulier les raisonnements scientifiques, et plus généralement d'interpréter les œuvres humaines à l'aide de techniques d'intelligence artificielle. Cette perspective ouvre naturellement sur les *humanités numériques*, ouverture qu'accroît encore la mise en place du Labex OBVIL.

Positionnement dans le contexte local/ national/ international : nous avons développé des outils originaux d'alignement, de cartographie des contenus et d'analyse stylistique que nous comptons développer et surtout tester en collaboration avec les équipes de littérature de la Sorbonne et en partenariat avec des éditeurs. C'est cette proximité avec des équipes de littérature et des éditeurs qui fait l'originalité de notre positionnement au plan national et international.

Objectifs de résultats à 5 ans : nous souhaiterions devenir l'une des équipes de référence dans le monde sur la thématique des humanités numériques. La conjoncture nous apparaît favorable pour deux raisons : nous possédons le savoir technique que nous avons déjà mis en œuvre dans différentes réalisations et grâce au Labex OBVIL nous entretenons des liens étroits avec les équipes de littérature de la Sorbonne. Nous nous fixons trois objectifs concrets d'ici à cinq ans. Le premier porte sur la réalisation d'outils nouveaux d'aide à l'interprétation des corpus textuels. Cela s'inscrit dans la lignée du logiciel MEDITE qui est un aligneur unilingue. Nous souhaitons l'étendre pour détecter les réemplois dans des corpus littéraires. Ce projet s'est amorcé dans le cadre du Labex OBVIL avec le PEPS PHOEBUS. Le second vise à la conception d'éditions électroniques nouvelles en collaboration avec des éditeurs. Cela doit en particulier conduire à s'impliquer dans la définition des standards, en particulier dans le format ePub. Le troisième porte sur la cartographie de contenus à l'aide d'îles de mémoire.

4.2 Thème 2 : traitement sémantique de l'information

Description : dans cette thématique, nous faisons appel aux approches symboliques de l'intelligence artificielle, en particulier aux graphes conceptuels, pour aborder le traitement sémantique de l'information. Cela nous a conduit à aborder la fusion symbolique d'information, l'étude de la typicalité à l'aide d'approches fondées sur la néo-topologie, l'induction automatique d'axiomes terminologiques, l'extraction d'information à partir de textes, la représentation d'ontologies et leur cartographie. Dans le futur, nous voudrions poursuivre dans ces directions et travailler aussi sur l'abduction, sur la *programmation logique abductive* et sur l'apprentissage collectif.

Originalité : les techniques de fusion symbolique et l'approche néo-topologique apparaissent originales. Nous souhaitons les valider à la fois au plan théorique et sur des applications réelles, par exemple, dans le domaine de la santé, en les couplant éventuellement à des traitements plus numériques pour extraire des représentations symboliques. De même, l'étude de la découverte scientifique, en particulier des processus

de découverte en médecine, avec des outils d'intelligence artificielle, par exemple avec des outils de modélisation de l'abduction, apparaît comme une piste originale.

Positionnement dans le contexte local/ national/ international : au plan international, de nombreuses équipes travaillent actuellement sur les logiques de description annulables (*Defeasible Description Logics*). Nous proposons une approche distincte que nous souhaitons valider au plan théorique. Quelques équipes travaillent sur l'abduction, en particulier en Angleterre et au Japon. Nous souhaitons poursuivre nos travaux en collaboration avec ces équipes.

Objectifs de résultats à 5 ans : valider les approches la de fusion symbolique sur des applications conséquentes, en particulier dans le domaine de la santé. Valider, au plan théorique les approches néo-topologiques. Préciser le rôle de l'abduction dans la découverte d'hypothèse et constituer un réseau international.

5 Analyse SWOT

Points forts : nous disposons d'un savoir technique important sur les humanités numériques et sur le traitement sémantique de l'information. De plus, le contexte, lié au Labex OBVIL, nous apparaît très favorable à la constitution d'un centre de rayonnement mondial sur cette thématique.

Points faibles : le point faible tient à la petite taille de notre équipe. Il faudrait pouvoir la renforcer. Nous le faisons en recrutant des chercheurs post-doctoraux sur contrats.

Risques liés au contexte : la crise économique risque de limiter les possibilités de recrutement. De plus, les financements du Labex promis sur 8 ans peuvent éventuellement se réduire.

Possibilités liées au contexte : outre le Labex OBVIL, nous devons reprendre et assurer le suivi des outils d'annotation sémantique développés par le professeur Jean-Pierre Desclés.

6 Mise en œuvre

Valorisation et recherche de moyens : deux projets investissement d'avenir se mettent en place, l'un avec la société Jouve et le groupe Editis, l'autre avec la société Actialuna et les éditions Viz. Plus généralement, nous souhaitons nous impliquer dans des programmes nationaux et internationaux portant sur les applications des humanités numériques pour l'édition électronique et la recommandation.

Diffusion des résultats (stratégie de publication) : nous souhaiterions renforcer nos publications en revue et dans les actes des grandes conférences internationales d'intelligence artificielle, d'humanités numériques et de philosophie, ce pourquoi la venue de Gauvain Bourgues devrait nous aider.

Stratégie locale / nationale/ Internationale : notre stratégie se déploie sur deux axes déterminés, l'un, par l'appartenance au Labex *OBVIL* dont l'équipe ACASA est l'un des porteurs, l'autre par les travaux sur le traitement sémantique de l'information. Au plan international, différents partenariats sont en cours de montage, en particulier avec les équipes d'humanités numériques du MIT, au travers du *Collège d'études mondiales* et de la fondation Mellon, avec des équipes européennes, à travers le groupe de réflexion "*On-Life*", enfin avec l'<idpf> (International Digital Publishing Forum), pour l'établissement de standards d'édition électronique.

Implication future dans l'enseignement : l'équipe ACASA s'implique à la fois dans les enseignements de master de l'UPMC, dans la mise en place d'enseignements internationaux de master et dans des programmes post-doctoraux internationaux.

- *Lien avec la nouvelle maquette du L et M* : participation active au montage de la spécialité DAC (Données, Apprentissage et Connaissances) du master.
- *Formation internationale* : Jean-Gabriel Ganascia est responsable du master Erasmus Mundus DMKM pour l'UPMC.
- *Échanges internationaux* : l'équipe ACASA participe au programme COFUND de financement de bourses post-doctorales internationales avec le *Collège d'études mondiales* dans le cadre d'un accord de partenariat entre le Labex OBVIL et la fondation de la Maison des Sciences de l'Homme.

Chapitre 6

Projet scientifique - Département NetSys

Sommaire

6.1	Projet 2014 – 2018 - Département NetSys	110
6.2	Projet 2014–2018 - Equipe PHARE	112
6.3	Projet 2014–2018 - Equipe NPA	115
6.4	Projet 2014–2018 - Equipe REGAL	119



Vague D : Campagne d'évaluation 2012-2013

6.1 Projet 2014 – 2018 - Département NetSys

1 Présentation du département

Le département *NetSys* (“*Networks and Systems*”) est issu de la séparation du département RSR en deux départements plus petits, le second département étant *Complex Systems*. *NetSys* réunit les activités de recherche du LIP6 liées aux réseaux, systèmes et systèmes répartis. Il se consacre à l’analyse et la conception de solutions pour construire et gérer les réseaux, systèmes et systèmes répartis du futur. Ces thèmes sont représentés par trois équipes complémentaires : *NPA*, *Phare* et *REGAL*, cette dernière étant aussi une équipe commune avec Inria Paris-Rocquencourt.

2 Périmètre scientifique

L’activité de recherche de *NetSys* se concentre sur l’ensemble des thématiques liées à la communication et aux systèmes informatiques, à savoir : le calcul distribué, l’Internet du futur, les systèmes pair à pair, les réseaux mobiles et ambiants, les systèmes large échelle, les nuages informatiques, et la tolérance aux fautes et aux attaques. Ces thèmes sont bordés à la fois sous des aspects théoriques et pratiques. Les méthodes utilisées par les trois équipes s’articulent principalement autour d’une ou plusieurs des approches suivantes :

- *Modélisation* : Il s’agit de formaliser certaines caractéristiques d’attaques, mobilité des équipements, capacités sensorielles, etc.),
- *Algorithmique distribuée* : Conception et mise en œuvre d’une algorithmique adaptés aux modèles, le cas échéant allant jusqu’à la mise en œuvre dans les systèmes et réseaux considérés,
- *Expérimentation* : Simulations, tests et mesures en grandeur réelle, prototypage, etc.

Les équipes vont principalement concentrer leurs recherches sur les sujets suivants :

- Algorithmes pour les environnements dynamiques
- Gestion distribuée des données dans les architectures multi-cœurs
- Gestion des contenus
- Réseaux sans-fil et mobilité
- Métrologie et gouvernance de l’Internet
- Accès hertzien de nouvelle génération
- Internet et Sky computing
- Intégration médicale dans l’Internet
- Internet et cybercriminalité

Par ailleurs, le département est fortement impliqué dans les quatre axes transverses du LIP6 suivants : *Fiabilité et sécurité*, *Traitement distribué de données distribuées* et *Nouveaux Systèmes Répartis*.

3 Positionnement

Le département réunit des chercheurs reconnus du domaine des systèmes, des systèmes distribués et des réseaux, tant du point de vue théorique que pratique. Les trois équipes constituant *NetSys* jouent un rôle central tant sur le plan national qu’international dans la compréhension des problèmes abordés et sont parfaitement positionnées.

4 Les défis scientifiques

Les défis scientifiques sont clairement liés aux mutations constatés dans l'utilisation des moyens informatiques. Désormais, les terminaux sont de plus en plus petits. Ils sont également mobiles, capables de percevoir leur environnement, etc. Cette évolution génère de nouveaux comportements, de nouvelles habitudes d'utilisations et de nouvelles possibilités d'applications, lesquelles doivent s'adapter en conséquence. A ces terminaux s'ajoutent des équipements autonomes tels que des capteurs, des robots mobiles, des véhicules, ainsi que des éléments réseaux eux-mêmes reliés à l'Internet. Ces appareils sont reliés entre eux via un réseau d'interconnexion ayant des caractéristiques différentes selon les éléments qu'il relie, les types de liaisons (filaire, optique, sans fil, etc), les contraintes environnementales et applicatives.

Les défis scientifiques tournent donc autour de ces points, à savoir :

- Concordance des modèles vis-à-vis des environnements réels,
- Adéquation des algorithmes vis-à-vis des problèmes à résoudre et des tâches correspondantes,
- Conception de solutions auto-organisées et capables de s'adapter aux changements d'échelle.

5 Evaluation du département : analyse SWOT

Points forts. Que ce soit au travers de projets formalisés ou non, les trois équipes travaillent en étroite collaboration sur les principales thématiques du département, notamment l'algorithmique distribuée, la tolérance aux fautes, l'Internet du futur et les systèmes distribués dynamiques à large échelle (systèmes pair à pair, réseaux mobiles hétérogènes, essaims de robots, etc.). Par ailleurs, les membres des trois équipes sont globalement fortement impliqués dans de grands projets de recherche, d'enseignements et industriels à tous les niveaux (local, national et international). Enfin, les trois équipes sont nationalement et internationalement reconnues et toutes ont d'étroites relations avec le monde industriel sous différentes formes (collaborations, start-up et programmes de formation).

Points faibles. Le département est issu de la scission en deux d'un département puissant par sa représentation au sein du LIP6 (en gros, un tiers des effectifs du laboratoire). Il faudra donc être vigilant pour maintenir une situation équilibrée notamment dans l'attribution des moyens pour *NetSys* et *Complex Systems* vis-à-vis des autres départements. Certaines équipes souffrent d'un manque d'encadrement (déséquilibre quantitatif entre chercheurs de rangs A et de rangs B).

Par ailleurs, si *NPA* et *REGAL* sont de taille équivalentes, *Phare* est quant à elle plus petite.

Risques liés au contexte. Du point de vue scientifique, les thèmes du département traitent de problèmes difficiles et très compétitifs. Certaines recherches risquent d'être reprises par des groupes plus importants, voire de ne pas aboutir à temps. Par ailleurs, on constate une baisse du nombre de projets s'appuyant sur des financements publics (ANR, région, etc.). Cette diminution s'avère être un risque important pour le département compte tenu de cette part importante de ses financements. Par ailleurs, la baisse constatée et programmée des dotations octroyées par les tutelles devrait également avoir un impact fort, aggravée par une politique de réduction des effectifs.

Possibilités liées au contexte. Le démarrage du *Labex SMART* au mois de juillet 2012 représente une perspective attractive et stimulante sur le moyen terme (jusqu'en 2020). *NetSys* se trouve être au cœur de l'un des quatre programmes du Labex. Ce programme est centré sur les environnements distribués conçus pour répondre à la mobilité des participants et à l'hétérogénéité des équipements. Ce thème est plutôt transversal à l'échelle du département car il met en œuvre un savoir-faire issu de plusieurs domaines, notamment les réseaux, les systèmes, les technologies du Cloud, la virtualisation et l'algorithmique distribuée. L'Equipex FIT constitue également une opportunité favorable.



Vague D : Campagne d'évaluation 2012-2013

6.2 Projet 2014–2018 - Equipe PHARE

1 Membres permanents de l'équipe PHARE

Nom	Prénom	Status	Etablissement
Kervella	Brigitte	MC	U. Amiens
Langar	Rami	MC	UPMC
Nguyen	Mai Trang	MC	UPMC
Pujolle	Guy	PU0	UPMC
Secci	Stefano	MC	UPMC

2 Evolution de l'équipe

La première évolution concerne la qualité de l'équipe qui fin 2014 devrait être composée de cinq Enseignants-Chercheurs "Habiletés à Diriger des Recherches". La survie de l'équipe PHARE proviendra du renouvellement du responsable et des Maitres de Conférences/HDR qui partiront vers des postes externes. Des arrivées permettant de préparer les remplacements seraient les bienvenues pour ne pas perdre les acquis et la reconnaissance internationale.

3 Caractérisation de la recherche

La caractéristique principale de l'équipe PHARE est d'aller de la conception au transfert technologique en passant par les différentes phases intermédiaires de modélisation, d'optimisation et de développement de maquettes. La conception peut passer par une étape de recherche fondamentale comme celle de la prochaine génération Post-IP à long terme. En effet, l'équipe PHARE reste attentive à cette nouvelle génération qui naîtra avec certitude à plus ou moins long terme.

4 Projet scientifique

L'équipe doit continuer dans sa lancée sur le développement de la nouvelle génération Internet mise au point pendant la période précédente par une plus forte intégration d'autres aspects importants de l'Internet. Les éléments clés restent la virtualisation, la très haute sécurité, les vitesses de transfert des ressources virtuelles et les accès sans fil et mobiles. Les éléments qui intéressent l'équipe dans les prochaines années correspondent à :

- l'Internet des choses avec les RFID et les équipements en nano et picotechnologie et leur intégration dans l'Internet, en particulier dans le cadre médical.
- La connexion hertzienne et son optimisation à tout moment dans le cadre de réseaux mesh multi-homés, multi-technologies, par radio cognitive et utilisant des le codage réseau et le contrôle de puissance.
- L'architecture Internet avec la nouvelle génération de Clouds virtualisés, les "Sky", avec une forte sécurité, un contrôle adéquat et une optimisation de l'algorithmique associée.
- La sécurité avec de nouveaux développements importants dans les cœurs de réseau pour contrer la cybercriminalité sur l'Internet.

Les sujets prospectifs de l'équipe PHARE sont également portés par des projets transverses à l'intérieur du LIP6 mais également à l'extérieur. L'équipe est actuellement impliquée dans deux axes transversaux du LIP6 : *objets communicants* et *optimisation*. Un sous-axe d'*objets communicants* concerne l'étude de technologies et d'algorithmes pour la radio cognitive, nécessitant à la fois une forte compétence en réseau et en traitement du signal. Sur ce sujet, les équipes PHARE et CIAN ont proposé un sujet de thèse commun qui a été retenu pour une allocation de thèse de l'école doctorale EDITE (2013-2016). L'équipe CIAN travaillera sur le niveau physique et l'équipe PHARE sur le niveau MAC pour les réseaux véhiculaires à radio cognitive. Une collaboration est également en cours avec l'équipe NPA sur la définition de techniques avancées de localisation en 3D des RFID.

L'équipe est également impliquée dans un des deux sous-axes du thème *optimisation*. L'intérêt de la coopération transverse porte principalement sur la découverte et la définition d'approches algorithmiques plus aptes à résoudre des problèmes d'optimisation intégrant des objectifs multiples, des contraintes de différentes natures ou un contexte de décision multi-décideur (théorie des jeux). Des synergies sont déjà établies avec l'équipe DECISION sur l'étude de problèmes de routage coordonné multi-critères.

A l'extérieur du laboratoire, de nombreuses collaborations sont en cours et prévues : avec POSTECH, l'U. Waterloo et l'U. Paris-Descartes sur l'internet des choses dans le cadre médical ; avec l'U. de Waterloo (à travers un projet PICS) et l'U. Miami sur les réseaux de femtocells ; avec l'INRS sur la radio cognitive et les femtocells ; avec Télécom ParisTech sur le codage réseau et le routage Internet ; avec l'U. Genève, la DCRN et l'ESIEA sur la sécurité et la cybercriminalité ; avec le MIT et Orange Labs sur mobilité et Cloud ; avec l'UFRJ et l'UQAM sur la virtualisation des réseaux ; avec l'U. Missouri Kansas City et GMU sur l'optimisation des datacenters et la résilience dans le Cloud. De nombreux échanges d'étudiants et chercheurs et le montage des projets de recherche avec ces partenaires sont déjà en cours et d'autres sont prévus à court terme.

4.1 Thème : l'intégration médicale dans l'Internet

L'Internet devra supporter des milliards de connexions avec des équipements orientés médical dans un moyen/long terme. Les équipements réseaux sur le corps ou introduits dans le corps seront virtualisés et devront permettre une haute disponibilité ainsi qu'une très forte sécurité. Des microcontrôleurs sécurisés, des techniques de localisation, des techniques de redondance sont à mettre en œuvre.

L'originalité ne provient pas du domaine lui-même mais de la façon de traiter le sujet par la virtualisation et la sécurité de l'environnement virtuel. L'utilisation de nano capteurs virtuels donne naissance à une nouvelle génération facilement intégrable à l'Internet.

La force de l'équipe provient d'une connaissance importante des RFID et de la nanotechnologie au travers du projet Spinnaker, d'une approche nouvelle de la virtualisation avec la possibilité de l'intégrer dans des objets de petite taille et de sa compétence forte sur les microcontrôleurs sécurisés. Le résultat à cinq ans est une forte extension de l'architecture avec la possibilité d'y intégrer des nano machines virtuelles personnalisées.

4.2 Thème : Accès hertzien de nouvelle génération

Le trafic sur les accès hertziens double toutes les années entre 2010 et 2020, ce qui correspond à une demande d'accès multipliée par 1 000. Pour atteindre ces capacités un besoin impératif d'innovations est nécessaire et proviendra de l'utilisation simultanée de technologie en cours de mise au point incluant la radio cognitive, des cellules de plus en plus petites (femtocell mais aussi attocell), du codage réseau et du contrôle de puissance. PHARE souhaite développer une nouvelle génération intégrant ces solutions.

L'originalité provient à la fois de la taille personnalisée des cellules pouvant aller de quelques mètres à quelques centaines de mètres et optimisée en temps réel pour satisfaire les contraintes des applications, tout ceci avec une utilisation d'approches mesh en radio cognitive avec contrôle de puissance et en utilisant du codage réseau dès que le maillage est suffisant.

La force de PHARE dans cette thématique provient de son actuelle compétence à la fois théorique et pratique de ces domaines. Les résultats dans cinq ans seraient l'intégration de l'ensemble des technologies décrites dans une nouvelle génération fondée sur la variabilité de la taille des cellules. Une des applications visées est un stade avec 50000 utilisateurs et un débit moyen de 1 Mbit/s par utilisateur.

4.3 Thème : Internet et Sky

Les Clouds virtualisés forment une nouvelle génération de Clouds capables de répondre simultanément à toute demande de tout type se trouvant sur n'importe quel Cloud du monde. Cette technologie qui a commencé à être introduite par l'équipe PHARE sous le nom de "Sky" est une combinaison par virtualisation de l'ensemble de ce que l'on peut trouver dans les Clouds d'aujourd'hui et de demain. L'opérateur d'un "Sky" devra pouvoir rapatrier des machines virtuelles de tout type se trouvant n'importe où dans le monde avec une haute sécurité et des vitesses de déplacement se comptant en Gbit/s. Les protocoles de localisation multi-niveaux comme LISP, TRILL et MPTCP devront faire appel à du multi-routage à grande échelle et à une sécurité fortement accrue par rapport à la version actuelle.

Le concept "Sky" a été lancé par PHARE en 2010, mais sans en dévoiler toutes les potentialités et l'importance de ce nouveau paradigme, et la force de PHARE est d'être déjà au cœur du sujet. L'équipe est en train de s'allier avec de nombreuses PME françaises pour réaliser cette nouvelle génération de "Sky". L'objectif est d'obtenir un contrat pour son développement dans le cadre des Investissements d'Avenir pour en avoir une réalisation complète et opérationnelle dans cinq ans.

4.4 Thème : Internet et cybercriminalité

Les cyberattaques sont devenues officiellement une nouvelle arme que les états se doivent de contrer. L'ambition est de construire les éléments nécessaires pour contrer ces attaques. La vision actuellement en cours de réflexion se fonde sur des microcontrôleurs sécurisés et une algorithmique importante dans les cœurs de réseaux et les clouds. Des accords avec différents centres de recherche sont en cours pour partager les connaissances de ce domaine complexe. L'objectif à cinq ans est de développer un environnement algorithmique fondé sur des microprocesseurs sécurisés capables d'arrêter la plupart des attaques de cybercriminalité.

5 Analyse SWOT

Points forts : très forte compétence des membres permanents de l'équipe avec une complémentarité en harmonie avec les recherches. Alliance très serrée (quoique non formalisée) avec des équipes de recherche de pointe reconnues internationalement et forte compétence industrielle au travers des start-up de l'équipe.

Points faibles : équipe légèrement trop petite pour progresser suffisamment rapidement sur des sujets majeurs. De nombreuses recherches sont reprises par des groupes plus importants.

Risques liés au contexte : : disparition de l'équipe PHARE à un terme de quelques années.

Possibilités liées au contexte : forte potentialité liée au contexte recherche-industrie développé par l'équipe.

6 Mise en œuvre

La mise en œuvre des futures recherches de l'équipe PHARE s'effectue en premier lieu par l'équipe elle-même qui est homogène et qui atteint tout juste la masse critique nécessaire pour asseoir sa renommée internationale. Ensuite, elle se fonde sur ses relations rapprochées avec des groupes solides et par sa surface industrielle au travers de ses start-up. Le groupe est actif dans le montage de projets avec le regret que les résultats soient assez aléatoires par rapport aux projets déposés, ce qui implique le dépôt d'un nombre de projets importants pour permettre la continuité des recherches.

La diffusion des résultats a toujours été un point fort aussi bien au niveau des revues que du passage au stade industriel. Cette stratégie devrait continuer par la publication dans de grandes revues internationales et par le maquettage pour s'assurer de la qualité des recherches et permettre le transfert technologique.

PHARE a fait un effort disproportionné par rapport à sa taille et ses objectifs pour s'inscrire dans des actions transverses du laboratoire et à l'international en espérant que ces actions apporteront des fruits à la hauteur des espoirs. L'équipe PHARE fait partie de deux actions transverses du Laboratoire : l'axe "Objets Communicants" et l'axe "Optimisation". La diffusion s'effectue par les grands congrès que ce soit ceux de l'équipe ou ceux de l'extérieur et par les réseaux tissés depuis de nombreuses années avec les meilleurs centres de recherche.

Les implications dans l'enseignement est une aspiration forte de l'équipe pour apporter sa compétence à la fois théorique et pratique du monde des réseaux. L'équipe PHARE est bien intégrée dans la nouvelle maquette de la Licence et de la Maîtrise.



Vague D : Campagne d'évaluation 2012-2013

6.3 Projet 2014–2018 - Equipe NPA

1 Membres permanent de l'équipe

Baey	Sébastien	MC	UPMC
Baron	Loïc	IE2	CNRS
Baynat	Bruno	MC0	UPMC
Blin	Lélia	MC	UNIV D'EVRY
Cruz Teixeira	Renata	CR1	CNRS
Dias de Amorim	Marcelo	DR2	CNRS
Fdida	Serge	PUEX	UPMC
Fladenmuller	Anne	MC	UPMC
Fourmaux	Olivier	MC	UPMC
Friedman	Timur	MC	UPMC
Malouch	Naceur	MC	UPMC
Marzouki	Meryem	CR1	CNRS
Pau	Giovanni	Chaire	ATOS / Renault
Potop-Butucaru	Maria	PU2	UPMC
Spathis	Prométhée	MC	UPMC
Thai	Kim Loan	MC	UPMC
Tixeuil	Sébastien	PU2	UPMC

2 Evolution de l'équipe

Le recrutement en septembre 2012 sur un poste de Professeur de Maria Potop-Butucaru devrait permettre de développer l'activité de l'équipe en algorithmique distribuée notamment en liaison avec les thématiques "Contenus" (distribution pair à pair) et "Internet du Futur" (météorologie distribuée). Le recrutement en fin d'année 2012 de Giovanni Pau sur une chaire industrielle financée par ATOS et Renault devrait donner l'impulsion nécessaire à l'établissement d'une activité visible autour des réseaux de véhicules et de la vie à bord et renforcer les thématiques "Réseaux dynamiques et spontanés" et "Contenus". Deux chargées de recherche et une MC sont HDR, et trois MC sont en passe de passer l'HDR à leur tour. Il est donc probable que des départs aient lieu, et il conviendra de recruter en conséquence.

3 Caractérisation de la recherche

La recherche menée dans l'équipe présente un spectre particulièrement large, puisque nous considérons les modèles et algorithmes répartis pour les réseaux dynamiques émergents, mais avons aussi une activité pratique qui va jusqu'à la mise en oeuvre de plates-formes d'expérimentations à échelle mondiale en passant par le développement et l'utilisation de nouvelles techniques de simulation.

4 Projet scientifique

Le projet scientifique de l'équipe peut être décliné suivant deux axes, celui des *réseaux* et celui des *usages*. Les réseaux du futur englobent le futur cœur de l'Internet mais aussi tous les réseaux émergents et omniprésents dont le dynamisme à de multiples niveaux remet en cause nombre de solutions déployées aujourd'hui. L'évolution des usages fait qu'aujourd'hui les réseaux sont essentiellement utilisés pour transférer du contenu (et de nouveaux paradigmes sont à définir pour répondre à ces besoins), mais la plus grande décentralisation du contrôle pour faciliter le déploiement et l'auto-organisation a rendu possible nombre de vulnérabilités qu'il convient de circonscrire.

4.1 Réseaux dynamiques et spontanés

Description L'Internet actuel suit un modèle où le réseau est au centre, dans lequel l'utilisateur subit des contraintes structurelles. Afin que l'Internet du futur puisse être construit par les utilisateurs et pour les utilisateurs, la recherche sur les réseaux de communication devrait se concentrer en particulier sur la "création de substrats pour que les utilisateurs puissent développer leurs propres Internets".

Originalité L'équipe adopte une méthodologie alliant théorie et pratique. L'originalité des travaux repose sur une équipe ayant des expertises très variées et complémentaires sur l'algorithmique, la modélisation et l'implémentation. L'équipe propose de continuer sa méthodologie de recherche en développant des activités sur la mesure, l'analyse et la compréhension de phénomènes émergents de l'intégration des utilisateurs (et de l'impact de leurs comportements spatiotemporels) dans la conception de systèmes de communication en réseau.

Positionnement dans le contexte local/ national/ international Au niveau national, les travaux menés sont régulièrement exposés dans les événements principaux de la communauté (Algotel, Rescom, CFIP) et figurent dans des projets à la fois industriels et académiques. Les membres de l'équipe sont souvent appelés pour des travaux d'expertise et pour des jurys (de thèse, de HDR, de sélection). Au niveau international, nous collaborons avec les meilleures équipes du monde et publions dans les meilleurs journaux et conférences du domaine. Par le biais de projets Européens principalement, l'équipe souhaite maintenir sa place d'acteur reconnu dans le domaine.

Objectifs de résultats à 5 ans Outre la poursuite de notre activité de modélisation et d'analyse de performances, nous voulons continuer nos travaux sur la conception et la mise en pratique de solutions pour les réseaux mobiles à connectivité intermittente. Une application qui nous intéresse en particulier est la décharge de trafic (*data offloading*) qui permettra d'alléger le trafic des réseaux cellulaires. Nous travaillerons sur les stratégies d'ingénierie de service dans ces réseaux en essayant de les expérimenter sur des plateformes réelles de radios traditionnelles et cognitives. Un projet Européen vient d'être accepté sur cette question, dont l'idée est de concevoir, dimensionner, implémenter et évaluer une nouvelle architecture de communication qui supporte des stratégies dynamiques de décharge de trafic. L'originalité de ce projet est le contrôle effectué par l'opérateur.

Sur les questions liées à l'accès au médium, nous proposerons des mécanismes d'ordonnancement entre les radios cognitives afin d'assurer un partage équitable de la bande passante. Le concept d'allocation de ressources opportuniste offre également des perspectives intéressantes dans un contexte semi-décentralisé voire complètement décentralisé.

Les 5 prochains ans seront aussi certainement marqués par une forte activité autour des réseaux véhiculaires. L'équipe, qui collabore déjà avec Toyota, accueillera début 2013 un chercheur spécialisé dans le domaine dans le cadre d'une chaire de l'UPMC avec Renault et Atos. Cette chaire, étalée sur 5 ans, a pour but de placer le LIP6 au plus haut niveau mondial sur les problèmes de "vie à bord".

4.2 Internet du Futur

Description Nous travaillons sur trois axes : (1) les mesures d'Internet, avec un accent sur les diagnostics de pannes, (2) le déploiement de plateformes d'expérimentation pour l'Internet du futur, fédérées à l'échelle planétaire et (3) la gouvernance d'Internet, en regardant de près le filtrage de flux et la protection des données personnelles.

Originalité En métrologie, l'instrumentation des machines des utilisateurs à des fins de diagnostic est une approche nouvelle avec des perspectives ouvertes. Concernant les plateformes, nous proposons des outils plus modulables et souples que ce qui existe déjà pour favoriser la fédération. En gouvernance, nous avons une approche d'identification croisée des méthodologies et des politiques.

Positionnement Nous sommes leaders en Europe et en France en matière de la fédération de plateformes d'expérimentation, avec la coordination de deux Projets Intégrés FP7 (OneLab2 et OpenLab) et d'un Equipement d'excellence français (FIT), et nous représentons souvent l'Europe en ce domaine dans des réunions internationales avec les Etats-Unis, le Japon, la Chine, la Corée du Sud, le Brésil, et d'autres pays.

Objectifs de résultats à 5 ans Nous cherchons à aller au delà de notre impact dans la communauté scientifique pour voir le déploiement d'outils basés sur nos idées. Un grand nombre d'utilisateurs d'Internet peuvent profiter de nos méthodes de diagnostic de pannes. Les plateformes qui sont aujourd'hui expérimentales peuvent devenir dans cinq années un noyau de l'éventuel Internet du Futur. Nous visons à ce que nos résultats sur la gouvernance d'Internet trouvent leur traduction sur un plan pratique au quotidien.

4.3 Tolérance aux pannes et aux attaques

Description Les réseaux actuels et futurs font la part belle à l'auto-organisation, la virtualisation, l'adaptabilité, le passage à l'échelle, et d'une manière générale à la distribution du contrôle. Une telle distribution induit une plus grande vulnérabilité aux fautes (dont la probabilité d'occurrence augmente) et aux attaques (plus d'utilisateurs et une grande hétérogénéité des matériels et logiciels utilisés).

Originalité Les approches développées dans l'équipe partent du principe que toute solution de sécurité (matérielle ou logicielle) peut être compromise qu'il est impossible de distinguer une machine mal initialisée mais bien intentionnée d'une machine dont l'intention de nuire est avérée. Cette approche permet en premier lieu de donner des indications sur les garanties qui peuvent être fournies en cas de compromission du réseau, mais aussi de connaître les limites de toutes les approches qui pourront être développées dans un contexte donné.

Positionnement dans le contexte local/ national/ international La plupart des travaux existants adressent un type de fautes ou d'attaques particulier. Notre expertise sur la multitolérance est reconnue par plusieurs équipes aux USA, au Japon et en Israël, avec qui nous publions régulièrement. Une montée en puissance au niveau européen sur ce domaine, par exemple via le montage de projets collaboratifs, est souhaitable.

Objectifs de résultats à 5 ans Nos objectifs sont ambitieux. Tout d'abord, nous souhaitons intégrer diverses notions de pannes "fréquentes" (va et vient des utilisateurs, mobilité, arrêt et redémarrage inopiné) avec des fautes et attaques plus malveillantes (attaques Sybilles, conspirations et coalitions, interférences volontaires et déni de service). En plus des résultats théoriques attendus, nous souhaitons développer une suite de logiciels qui permette les benchmarks de tolérance aux pannes et aux attaques, que ce soit au niveau de la simulation que de la plate-forme réelle.

4.4 Contenus

Description Les contenus sont aujourd'hui omniprésents comme l'attestent la multiplication des centres de données ou le succès des plateformes de "cloud computing". Pour améliorer la qualité de l'expérience utilisateur, il est nécessaire de concilier leurs actions avec les politiques des opérateurs de réseaux qui visent à optimiser l'utilisation de leurs ressources.

Originalité L'originalité des travaux repose sur des expertises variées et complémentaires sur la mesure des trafics réseau des applications émergentes, l'analyse des usages et la conception de nouvelles architectures de communication pour la distribution de contenu.

Positionnement dans le contexte local/ national/ international De nombreuses équipes de renom, tant académiques qu'industrielles, travaillent sur ces contenus qui constituent aujourd'hui l'essentiel du trafic réseau. Les travaux de l'équipe sont connus et appréciés au national comme à l'international.

Objectifs de résultats à 5 ans Avec l'augmentation massive de la diffusion de contenu, les besoins des fournisseurs, des opérateurs et des utilisateurs de contenus seront difficile à satisfaire conjointement. Nous

proposons de poursuivre l'effort vers plus d'auto-organisation de ces réseaux basés sur les flux d'information véhiculés. En particulier, nous souhaitons mettre l'accent sur la tolérance à la dynamique (induite par les reconfigurations du réseau sous-jacent, les arrivées et départ intempestifs de utilisateurs, mais aussi le comportement des utilisateurs lui-même) et au confinement de comportements malveillants dans les infrastructures de communication.

Nous souhaitons également porter une attention particulière aux contenus vidéo indépendamment des architectures supports (CDN, P2P, Hybrides...).

5 Analyse SWOT

Points forts : L'équipe possède une assise internationale attestée par le pilotage de plusieurs projets d'envergure. Le large spectre de compétences rend possible l'interaction entre les aspects théoriques et pratiques.

Points faibles : Les verrous scientifiques étant jugés importants, le domaine est très compétitif.

Risques liés au contexte : Le projet est ambitieux. La recherche de moyen et la politique de recrutement de l'équipe doivent être en phase avec ces ambitions.

Possibilités liées au contexte : En cas de succès l'impact potentiel est important tant du point de vue théorique que pratique.

6 Mise en œuvre

Valorisation et recherche de moyens : Nous avons une activité de collaboration nationale et internationale particulièrement active, tant sur le plan académique que sur le plan industriel. Nous comptons évidemment poursuivre dans cette voie, et soumettre des propositions pour des projets très visibles à l'international, comme les différents appels de l'ERC.

Diffusion des résultats (stratégie de publication) : Nous comptons continuer à publier dans les meilleures revues et conférences internationales, tout en préservant l'activité autour de sujets émergents lors de conférences et ateliers plus spécialisés.

Stratégie locale / nationale/ Internationale : L'équipe est impliquée dans le labex SMART, co-dirige le laboratoire LINCOS en partenariat avec Inria, Telecom ParisTech, et Alcatel-Lucent, et participe au pôle RESCOM du GDR ASR. Nous coordonnons également plusieurs projets européens et avons de nombreuses action bilatérales suivies avec des universités de renom.

Implication future dans l'enseignement : L'équipe est partie prenante dans le KIC ICT Labs, essentiellement au travers de son action de coordination nationale sur les Master (ICT Labs Master School) et coordination européenne de la Doctorate School. Dans la maquette d'habilitation du nouveau master de la spécialité "Réseaux", l'équipe a proposé d'assurer la responsabilité de 14 nouvelles UEs.



Vague D : Campagne d'évaluation 2012-2013

6.4 Projet 2014–2018 - Equipe REGAL

1 Membres permanents de l'équipe

Nom	Prénom	Statut	Etablissement
Arantes	Luciana	MC	UPMC
Darche	Philippe	MC	IUT PARIS 5
Folliot	Bertil	PU1	UPMC
Lawall	Julia	DR2	INRIA
Makpangou	Mesaac	CR	INRIA
Marin	Olivier	MC	UPMC
Monnet	Sébastien	MC	UPMC
Muller	Gilles	DR2	INRIA
Petit	Franck	PU1	UPMC
Sens	Pierre	PU1	UPMC
Shapiro	Marc	DR1	INRIA
Sopena	Julien	MC	UPMC
Thomas	Gaël	MC (HDR - Sept. 2012)	UPMC

2 Evolution de l'équipe

L'équipe possède au début 2012 un bon équilibre des permanents entre les aspects algorithmiques et systèmes. Cependant, la composition de l'équipe est amenée à profondément évoluer sur les prochaines années. Il y a tout d'abord le départ au 1^{er} septembre 2012 de Maria Potop-Butucaru qui travaillait sur les algorithmes et qui a été nommé professeur. De plus, trois maîtres de conférences de l'équipes vont soutenir fin 2012 début 2013 leur HDR. Nous prévoyons donc dans les 6 années de nouveaux recrutements pour maintenir l'équilibre entre les aspects théoriques et pratiques qui font l'originalité de l'équipe.

3 Caractérisation de la recherche

Notre recherche est à la fois de nature fondamentale, avec la proposition de nouveaux modèles et algorithmes dans les environnements répartis émergents, mais aussi appliquée avec le développement de prototypes et de logiciels. Ce double positionnement est original et permet une validation pratique de nos résultats théoriques.

4 Projet scientifique

Notre projet scientifique s'articule autour des deux thèmes qui correspondent à des évolutions majeures des infrastructures réparties.

Les grandes infrastructures réparties sont de plus en plus dématérialisées et dynamiques : dans les nuages informatiques le lien entre les nœuds logiques (les machines virtuelles) et les machines physiques n'est plus figé et peut évoluer au cours du temps en fonction des variations de charge. De même, dans les réseaux

dynamiques (réseaux de capteurs, réseaux tolérant les délais ou *DTN*...) les périphériques interagissant sont mobiles et leurs connexions intermittentes. Même si ces environnements apparaissent éloignés, les deux reposent sur une topologie (logique dans le cadre des clouds et physique dans le cadre des réseaux) dynamique. Prendre en compte la dynamique de ces topologies est un défi important pour la communauté des chercheurs en algorithmes répartis dans lequel peu de résultats existent. Le défi consiste à proposer des modèles de systèmes répartis intégrant la dynamique au-dessus duquel une nouvelle algorithmique pourrait être construite.

Le nombre de cœurs de machines augmentent : Le nombre de cœurs (unités de calcul) intégrés au sein d'un microprocesseur ne cesse de croître, par exemple le processeur expérimental d'Intel Polaris contient d'ores et déjà 80 cœurs. Contrairement aux architectures distribuées en "ferme de serveurs", ces nouveaux multi-cœurs présentent l'avantage d'offrir une mémoire partagée. Cependant, les architectures exhibent des accès mémoire non uniforme (NUMA), c'est-à-dire que la latence d'accès aux données dépend à la fois de l'adresse et du cœur qui y accède. On constate que les applications actuelles sont incapables de tirer parti de la puissance de calcul de ces architectures car les systèmes existants sont conçus avec une hypothèse d'uniformité d'accès. Le défi concerne l'optimisation des applications pour les architectures NUMA, en gérant le placement mémoire au sein des systèmes et des machines virtuelles langages.

Notre projet de recherche sur les prochaines années aborde ces deux thèmes majeurs.

4.1 Algorithmes pour les environnements dynamiques

Dans ce thème nous envisageons deux voies de recherche respectivement sur les modèles et les algorithmes.

Description. *Modèles des systèmes dynamiques* : En algorithmique, les modèles de systèmes permettent de caractériser formellement les propriétés de l'architecture et des applications : type de fautes, délai de communication modèle de l'adversaire, graphe de communication etc. Ces modèles sont utilisés ensuite pour exprimer sur les briques algorithmique de base des systèmes (consensus, l'élection ou l'exclusion mutuelle) des résultats d'impossibilité, de faisabilité ou de minimalité. Or la plupart de ces résultats ont été obtenus à partir de modèles de systèmes statiques. Un de défis auquel est confronté notre communauté est de définir un modèle théorique permettant de caractériser la dynamique des systèmes et notamment de prendre en compte l'évolution dans le temps du graphe de connexion. Dans ce contexte, REGAL compte contribuer en définissant un modèle adapté aux réseaux hybrides composant de machines fixes et mobiles. Ce modèle permettra notamment d'exprimer des conditions pour résoudre des problèmes fondamentaux comme le consensus ou la dissémination fiable d'informations.

Algorithme en présence d'incertitude : Dans les environnement dynamiques, la forte volatilité des nœuds interdit une vision globale du système et implique des prises de décisions en fonction de visions très partielles voir fausses de l'environnement. Il faut donc définir une algorithmique tolérant les incertitudes sur les participants. Deux voies nous semblent prometteuses : l'algorithmique indulgente où l'incertitude est un paramètre d'entrée et les algorithmes probabilistes. Dans ce cadre, nous avons commencé à explorer les algorithmes probabilistes de dissémination d'informations assurant un certain taux de couverture (ce travail préliminaire fait l'objet d'une publication à SRDS 2012).

Positionnement dans le contexte local/ national/ international. Il existe peu de modèles qui prennent en compte la dynamique. Des modèles comme celui des "infinite arrivals" [4] proposé par l'équipe de Roberto Baldoni (Université de Sapienza de Rome) considère l'arrivée de nouveaux nœuds ainsi que des évolutions dans les liens de communication. En revanche, ils maintiennent des hypothèses de connectivité forte entre les nœuds (pas de partitions) peu compatibles avec les environnements hautement dynamiques. Récemment, l'équipe de Nicolas Santoro (Université d'Ottawa) a proposé un modèle élégant de graphes dynamiques appelé TVG (Time varying Graph) [3] inspiré des "evolving graph". Ce modèle intègre la notion de chemin dans le temps. À partir des TVG, nous souhaitons proposer des modèles adaptés à des réseaux partitionables et sujets aux défaillances. Cette thématique est au cœur du 5 thèmes scientifique "Automatic Distributed Environments For Mobility" du labex SMART. Au sein du labex l'équipe assure la coordination scientifique de cette thématique.

Objectifs de résultats à 5 ans. Notre objectif est (1) de proposer un modèle inspiré des graphes dynamiques qui permette de modéliser la mobilité et les fautes, (2) de concevoir des algorithmes qui supportent une incertitude sur la vision de l'environnement.

4.2 Gestion des données dans les architectures à grand nombre de cœurs

La communauté scientifique a identifié deux verrous scientifiques importants liés aux environnements multi-cœurs sur lesquels l'équipe va se concentrer dans les années à venir : sur la conception du Ramasse-Miettes, et l'optimisation des algorithmes de synchronisation.

Description. *Ramasse-Miettes* : Les langages de programmation modernes s'appuient sur un mécanisme de gestion automatique de la mémoire, dit ramasse-miettes(RM). Or, nos premières expérimentations ont montré que le RM devient un goulot d'étranglement à partir de 10 à 12 cœurs. La cause première est lié au NUMA, car le RM fait beaucoup d'accès non-locaux. Pour passer à l'échelle au-delà de la centaine de cœurs, il faudra sans doute restructurer la machine virtuelle langage dans le sens d'un véritable système réparti.

Synchronisation : Le passage à l'échelle des applications sur les architectures multi-cœurs est aussi limité par les performances des algorithmes de verrouillage. La contention d'accès à la ligne de cache sur laquelle se trouve le verrou et le coût dû au transfert inter-cœur des données partagées en sont les principales raisons. Nous proposons d'explorer une nouvelle approche consistant à sacrifier l'un des cœurs pour exécuter les sections critiques et de remplacer les prises des verrous les plus utilisés par des transferts de contrôle vers ce cœur. Cette approche devrait permettre de limiter l'écroulement des performances lorsque de nombreux fils d'exécution essayent d'entrer simultanément en section critique et permettre d'améliorer la localité d'accès mémoire des applications.

Positionnement dans le contexte local/ national/ international. Contrairement aux architectures à plusieurs dizaines de millions de cœurs, comme dans le projet Exascale, requérant une réécriture complète des applications, le problème posé par les architectures à quelques dizaines de cœurs est de savoir comment exploiter cette puissance de calcul sans modifier les applications existantes afin d'éviter des coûts de développement colossaux. Depuis cinq ans, de nombreux travaux de recherche s'intéressent à permettre aux systèmes d'exploitation de tirer parti de ces architectures [1, 2]. Ces travaux améliorent uniquement les performances des systèmes d'exploitation mais ne permettent pas aux applications existantes qui tournent au-dessus de mieux tirer parti de ces architectures. Pour résoudre ce problème, deux axes ont été identifiés par la communauté scientifique depuis deux ans : le ramasse-miettes et les algorithmes de verrouillage. L'équipe Regal est bien positionné sur les deux sujets avec déjà deux publications de travaux en cours sur les deux sujets, chacune des publication réunissant des compétences en algorithmique distribués et en système d'exploitation.

Objectifs de résultats à 5 ans. Notre objectif est de définir un nouveau substrat de machine virtuelle optimisant les accès mémoires en abordant les deux principaux verrous qui limitent actuellement les performances des architecture multi-cœurs.

5 Analyse SWOT

Points forts : Notre projet aborde des problèmes et verrous clairement identifiés par la communauté. Nous nous focalisons sur des thèmes précis (un modèle de graphe dynamique, les techniques de ramasse miette et les algorithmes de synchronisation) sur lesquels nous avons déjà des résultats préliminaires encourageants publiés en 2012 dans des conférences de premier plan (Usenix ATC, ASPLOS, SRDS). Nos derniers projets collaboratifs (ANR MyCloud et ANR I-JVM) commencent à aborder ces thématiques.

Points faibles : Les verrous scientifiques étant jugés importants, le domaine est très compétitif.

Risques liés au contexte : Le projet est ambitieux. La recherche de moyen et la politique de recrutement de l'équipe doivent être en phase avec ces ambitions.

Possibilités liées au contexte : En cas de succès l'impact potentiel est important tant du point de vue théorique que pratique.

6 Mise en œuvre

Valorisation et recherche de moyens : L'algorithmique pour les environnements dynamiques dans le cadre des Nuages informatiques est déjà présent dans le projet ANR MyCloud auquel nous participons depuis 2011. Nous envisageons en 2013 la réponse à des appels d'offre des projets collaboratifs sur les aspects

plus fondamentaux. Sur l'axe multi-cœur, le nouveau projet ANR I-JVM que nous coordonnons porte sur cette thématique.

Diffusion des résultats (stratégie de publication) : Nos études préliminaires ont déjà fait l'objet de 3 publications en 2012 dans un journal international (The Computer Journal) et deux conférences de rang A (Usenix et SRDS). Nous envisageons de poursuivre nos efforts pour publier dans des conférences majeures.

Stratégie locale / nationale/ Internationale : L'axe sur les systèmes dynamiques s'inscrit parfaitement dans le cadre du Labex SMART porté par l'UPMC. Notre projet sur le multi-cœur est supporté par l'UPMC au sein d'un projet interne "Emergence".

Implication future dans l'enseignement : Dans la nouvelle habilitation de master d'informatique, une UE de M2 (NMV : Nouvelle Machine Virtuelle) étudiera spécifiquement les problèmes liés aux supports multi-cœurs. L'algorithmique pour environnements dynamiques sera étudiée dans l'UE ARA (Algorithmes Répartis Avancés) de M2.

7 Références

- [1] A. Baumann, P. Barham, P.-É. Dagand, T. L. Harris, R. Isaacs, S. Peter, T. Roscoe, A. Schüpbach, and A. Singhanian. The multikernel: a new os architecture for scalable multicore systems. In *22nd ACM Symposium on Operating Systems Principles (SOSP)*, pages 29–44, 2009.
- [2] S. Boyd-Wickizer, A. T. Clements, Y. Mao, A. Pesterev, M. F. Kaashoek, R. Morris, and N. Zeldovich. An analysis of linux scalability to many cores. In *9th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, pages 1–16, 2010.
- [3] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro. Time-varying graphs and dynamic networks. Technical report, University of Ottawa, 2011.
- [4] S. T. Piergiovanni and R. Baldoni. Eventual leader election in infinite arrival message-passing system model with bounded concurrency. In *Eighth European Dependable Computing Conference, EDCC-8*, pages 127–134, 2010.

Chapitre 7

Projet scientifique - Département SYSCOMP

Sommaire

7.1	Projet 2014 – 2018 - Département SysComp	124
7.2	Projet 2014–2018 - Equipe APR	128
7.3	Projet 2014–2018 - Equipe ComplexNetwork	132
7.4	Projet 2014–2018 - Equipe MOVE	135



Vague D : Campagne d'évaluation 2012-2013

7.1 Projet 2014 – 2018 - Département SysComp

1 Présentation du département

Le département *Complex Systems* est issu de la restructuration du département RSR en deux départements plus petits, l'autre département étant *NetSys (Networks and Systems)*. Il réunit trois équipes souhaitant mettre leurs forces en commun : APR (aspects méthodologiques basés sur l'algorithmique et la programmation), *Complex Networks* (aspects structurels avec comme formalisme central les graphes ou réseaux) et MoVe (étude, conception, sûreté et vérification des logiciels complexes). Les activités de chacune sont détaillées dans leurs fiches respectives.

2 Périmètre scientifique

Il y a système complexe dès lors qu'on fait face à un objet dont on n'arrive pas à décrire le comportement global par la connaissance, même fine, des éléments le constituant et de leurs interactions. Les exemples typiques sont les organismes vivants (et leurs écosystèmes), les systèmes industriels complexes (comme une usine ou un avion), ou les systèmes sociaux (comme un réseau social ou une ville, par exemple). En informatique, les exemples types sont les grands logiciels (assemblages complexes de briques logicielles), les réseaux (tels l'internet ou les réseaux pair-à-pair), et les usages qui en sont faits (réseaux sociaux en ligne, échanges, etc.).

Dans tous ces cas, l'expérience montre qu'analyser les composantes puis leurs interactions locales ne suffit pas à expliquer les comportements globaux des systèmes observés : on parle alors d'*émergence*. De même, concevoir des systèmes complexes fiables de grande taille devient un défi car il est délicat de maîtriser toutes les interactions de leurs composants. Dans cette situation, le tout est plus que la simple union des parties, et il s'agit de développer de nouvelles méthodes pour étudier, comprendre, concevoir et maîtriser ces systèmes.

3 Positionnement

Le département s'intéresse aux systèmes composés de grands ensembles d'entités en relation les unes avec les autres et pour lesquels la connaissance fine des entités et de leurs relations ne suffit pas, dans l'état actuel des connaissances, à comprendre l'ensemble du système. On considère en particulier des cas où le système a des propriétés globales inexplicables à partir des seules propriétés des entités le composant et leurs relations. En informatique, les exemples types sont les grands logiciels (assemblages complexes de briques logicielles), les réseaux (tels l'internet ou les réseaux pair-à-pair), et les usages qui en sont faits (réseaux sociaux en ligne, échanges, etc.). La recherche menée au sein du département vise à observer, décrire, comprendre, prédire et/ou contrôler de tels systèmes et leur évolution.

Il existe essentiellement deux façons de mener un travail sur un système complexe : une approche *top-down*, dans laquelle on considère le système comme un tout qu'on étudie en le divisant en parties de plus en plus finement, et l'approche *bottom-up*, dans laquelle on part des constituants et de leurs relations, pour arriver au système complet. L'exemple type est celui de l'internet : on peut le voir comme un objet global observé au travers de mesures et d'analyses de ses comportements, à la manière d'un organisme vivant, ou comme un assemblage de machines et de protocoles dont les spécifications, conçues par l'homme, devraient suffire à comprendre l'ensemble.

Le propre des systèmes complexes tels que nous les envisageons est justement le fait qu'aucune de ces approches n'est totalement satisfaisante. En effet, on n'arrive pas, dans la plupart des cas, à comprendre l'objet dans son ensemble à partir d'une connaissance fine de ses constituants ; réciproquement, on n'arrive pas à décomposer la plupart des systèmes complexes suffisamment pour les comprendre.

Notre positionnement dans ce contexte se veut profondément fondamental, dans le sens où il vise d'abord à *comprendre* et développer des *méthodes* pour les systèmes complexes. Afin d'atteindre cet objectif, et devant les insuffisances des approches actuellement disponibles, nous ne nous refusons par principe aucun moyen : nous sommes ouverts tant aux études empiriques qu'aux expérimentations, aux simulations, ou aux approches formelles. Notre conviction est que, au moins dans l'état actuel des connaissances, les progrès les plus importants viendront de travaux combinant ces méthodes.

Positionnement international. La thématique des systèmes complexes connaît un essor important depuis une dizaine d'années au niveau international. Elle s'est structurée, à l'international comme en France, en centres pionniers sur ces thèmes qui se sont maintenant organisés en groupes en assurant la cohérence, la promotion, l'animation et la coordination, comme le RNSC le fait au niveau français.

Au niveau européen, la *Complex Systems Society (CSS)* joue un rôle clé. Elle organise la *European Conference on Complex Systems (ECCS)* et coordonne la rédaction de *feuilles de route* pour la recherche en systèmes complexes. Nous sommes aujourd'hui très présents dans ces activités (membres de la CSS, publication de bulletins dans sa *newsletter ASSYST/CSS*, articles et workshops ECCS, préparation des appels d'offre européens tels ERA-NET CHIST-ERA, etc.). Nous restons toutefois en marge de certaines activités, comme la rédaction de la feuille de route européenne, et avons pour ambition de nous y impliquer davantage. En particulier, les thématiques informatiques n'y sont que très peu présentes, alors qu'elles ont un rôle crucial à jouer. Nous entendons leur donner toute leur dimension dans ce contexte.

Toujours au niveau européen, nous avons des relations étroites avec la plupart des institutions clés dans le domaine des systèmes complexes avec une composante informatique. Nommons pas exemple l'Université Catholique de Louvain (Belgique), l'*ISI Foundation* (Italie), l'Université de Milan (Italie), l'Université de Lancaster (Angleterre), ou l'Université Polytechnique de Catalogne. Nous avons des collaborations régulières avec ces institutions au travers d'invitations réciproques, de projets européens et de recrutements croisés.

Au-delà du niveau européen, nous sommes en relation avec les grands centres américains, asiatiques et africains sur le sujet, comme le Santa Fe Institute (USA), Indiana University (USA), Cornell University (USA), MAWILab (Japon), l'IIT de Kharagpur (Inde) ou l'université de Ouagadougou (Burkina-Faso). Ces relations nécessitent toutefois d'être renforcées, et nous planifions leur développement au travers d'invitations de chercheurs phares du domaine tels que Cris Moore (SFI), Mark Newmann (SFI), Jon Kleinberg (Cornell), Niloy Ganguli (Kharagpur) ou Alessandro Vespignani (Indiana) et l'échange d'étudiants en thèse ou postdoc.

Pour finir, nous avons pour ambition de publier certains de nos résultats les plus marquants dans les revues interdisciplinaires extrêmement prestigieuses *Science*, *Nature* et *PNAS*. En effet, sur le thème très transversal des systèmes complexes, ces journaux sont les endroits privilégiés pour donner à un résultat une portée et une visibilité extrêmement fortes. Or aujourd'hui les thématiques informatiques y sont très peu représentées, pour des raisons de positionnement et des raisons historiques. Nous comptons mobiliser nos efforts pour obtenir des publications de ce type, en plus des publications de premier rang en informatique que nous faisons déjà, ce qui donnerait une nouvelle dimension à l'impact de nos travaux.

4 Les défis scientifiques

Malgré un historique riche, les systèmes complexes restent un domaine largement en friche soulevant de nombreux défis.

Le plus fondamental d'entre eux est sans doute la convergence des approches : faute de méthodes générales et de succès satisfaisants des approches *top-down* et *bottom-up*, beaucoup de travaux reposent aujourd'hui sur des choix méthodologiques relativement arbitraires. Or il ne fait aucun doute que la vérité est dans la combinaison des approches, dans leur convergence : on devrait pouvoir à terme aborder un système complexe aussi bien du haut que du bas, puis mener l'étude jusqu'au bout et ainsi atteindre l'autre extrême. Nous pensons que cet objectif est le principal défi méthodologique à long terme du domaine.

A plus court terme (quelques années) et de façon plus pragmatique, nous identifions les défis suivants, sur lesquels nous comptons concentrer nos efforts.

Modélisation basée sur les mesures. Par nature, les systèmes complexes réels se comportent souvent différemment de ce que l'intuition aurait indiqué. Cela se vérifie par la mesure : l'observation des données collectées contredit l'intuition qu'on pouvait avoir concernant le comportement d'un système complexe. Or les modèles reposent généralement sur des formalisations de telles intuitions. Il est donc crucial dans notre contexte de développer une approche de modélisation *basée sur les mesures*. Un tel effort est déjà mené dans le cas de l'internet, où on a constaté à la fin des années 90 que la plupart des intuitions sur son comportement étaient en désaccord avec les mesures, mais il est loin d'avoir abouti et d'être général. Bien sûr, cette approche pose la question de la fiabilité de la mesure, qui peut être biaisée ou erronée. Il s'agit donc pour nous d'être capables de mettre au point et déployer des procédés de mesure capables de donner une information fiable sur les systèmes complexes concernés, qui sera confrontée aux modèles et susceptible de les invalider et/ou de guider le choix de modèles les plus pertinents. En allant plus loin, on demandera à ce que les modèles soient conçus pour tenir compte de ces propriétés observées en pratique.

Vers plus de formalisation. Dans beaucoup de cas, les études de systèmes complexes restent très empiriques : qu'elles reposent sur la mesure d'un système réel ou sur la simulation, on se contente souvent d'observer des propriétés et d'en dériver des conclusions. Ces travaux ne manquent pas d'intérêt et de rigueur, mais il est crucial pour le domaine d'arriver à en augmenter la formalisation. Ceci permettrait de déplacer le focus de la recherche du domaine vers des démonstrations rigoureuses de théorèmes plus généraux, et d'entamer une vraie compréhension des objets considérés. Bien sûr, cela suppose le développement de modèles appropriés (voir ci-dessus) et de méthodes pour les manipuler formellement. Nous pensons que les approches issues de la recherche en modélisation et en programmation pourraient aider grandement. Nous pensons également qu'il y a un fort besoin d'identification et de formalisation des problèmes soulevés par les systèmes complexes, afin que ceux-ci puissent éventuellement être abordés et résolus par une communauté de collègues plus large.

Gérer les grandes tailles. Les systèmes complexes sont souvent associés à des objets de grande taille, que ce soient les systèmes eux-mêmes ou les configurations qu'ils peuvent prendre, leur évolution temporelle, etc. Manipuler, stocker et traiter de telles tailles de données soulèvent de nombreuses questions algorithmiques difficiles. Typiquement, les complexités au-delà de quadratique sont proscrites, et on doit souvent se limiter à des complexités linéaires. Par ailleurs, le goulot d'étranglement pour les calculs sur les systèmes complexes se situe souvent au niveau de la complexité *en espace*. Celle-ci est trop rarement prise en compte dans les travaux algorithmiques, et nous souhaitons développer cette direction suivant deux axes : la maîtrise du coût en espace des calculs et le développement de techniques de représentations compactes pour pouvoir manipuler les données en mémoire. Cela est complémentaire d'approches telles que l'algorithmique de flots (la donnée arrive petit à petit et n'est pas stockée), qui nous semble également très prometteuse.

Dynamiques et hétérogénéités. Les objets que nous considérons sont souvent dynamiques (ils évoluent au cours du temps) et hétérogènes. Être capable de mesurer de telles dynamiques et hétérogénéités est un défi en soi ; les décrire, les caractériser, les capturer dans des modèles et des formalismes est encore plus difficile, mais nécessaire. Ces caractéristiques démultiplient les tailles des données considérées, et posent de nouvelles questions : dans la plupart des cas, les notions nécessaires pour y répondre restent à inventer.

Ces défis ont été identifiés comme extrêmement importants par l'ensemble des communautés travaillant sur les systèmes complexes. Ils réclament le développement de travaux autour de la mesure des systèmes déjà déployés (typiquement, cartographie de l'internet), la métrologie (biais induit, représentativité, etc.), l'analyse (description de ces systèmes), la modélisation (capture de leurs propriétés dans des modèles en permettant un traitement formel), et l'algorithmique (calcul efficace sur ces objets et représentation compacte). Les cas dynamiques et/ou hétérogènes, qui sont en fait prévalents, compliquent encore ces problèmes. Y faire face constitue le principal défi du département.

5 Evaluation du département : analyse SWOT

Points forts. Le département est un acteur clé au niveau national (hébergement du *Réseau National des Systèmes Complexes (RNSC)*, très impliqué dans les instituts des systèmes complexes parisiens et rhône-alpins (ISC-PIF et IXXI) ainsi que dans la rédaction de la feuille de route française pour les systèmes complexes). Au niveau international, il est un membre actif de la *Complex Systems Society (CSS)*, publiant régulièrement dans la conférence ECCS et dans le journal *Complex Systems*. Sa reconnaissance internationale est renforcée par son fort positionnement interdisciplinaire : publications de premier rang dans des domaines variés de l'in-

formatique (algorithmique, réseaux, vérification, etc.) mais aussi dans plusieurs autres disciplines (sociologie, physique, mathématiques). Enfin, il est constitué d'équipes très complémentaires, chacune avec une expertise de premier plan dans son domaine, et ayant la volonté de mettre leurs efforts en commun pour obtenir des avancées décisives dans le domaine des systèmes complexes.

Points faibles. Le département est en pleine création et son positionnement scientifique est en rupture avec les sujets traditionnels de l'informatique, sur un domaine nouveau et encore mal défini. Il s'agit donc pour nous de rapidement trouver, développer, exploiter et valoriser les synergies entre les équipes du département, ainsi qu'entre le département et les autres acteurs nationaux et internationaux. Il est notamment crucial pour le département de consolider sa place *à la fois* dans le paysage des systèmes complexes *et* celui de l'informatique, notre domaine d'ancrage (voir ci-dessous). Par ailleurs, il doit être bien clair que les défis scientifiques à relever dans notre domaine sont extrêmement ambitieux, et que le département a certainement besoin de cibler des priorités pour espérer y contribuer de façon majeure.

Risques liés au contexte. La principale difficulté liée à notre contexte est certainement liée au double positionnement *systèmes complexes* et *informatique* : le défi est d'être à la fois reconnu dans la communauté de recherche en systèmes complexes, dans laquelle les chercheurs en informatique sont très minoritaires, et dans la communauté de recherche en informatique dont nous sommes issus et qui restera résolument notre domaine d'ancrage. Ceci est en effet crucial pour de nombreux points clés, comme l'évaluation de notre production scientifique (notamment les publications) et de notre activité en général, la possibilité de recruter des candidats de première qualité (obtention de postes et attractivité) ainsi que pour le devenir de nos doctorants (qui pour être recrutés ont besoin d'un positionnement disciplinaire clair). Nous devons pour cela veiller à l'excellence de nos productions et, sans perdre notre composante interdisciplinaire, veiller à notre pleine reconnaissance dans le domaine de l'informatique.

Possibilités liées au contexte. Les tissus locaux (LIP6, UPMC, ISC-PIF, etc.), nationaux (CNRS, RNSC, IXXI, etc.) et internationaux (émergence de nombreux groupes sur la thématique) sont extrêmement favorables. Il existe notamment à l'UPMC le Programme Doctoral International *Systèmes Complexes*, le laboratoire *Unité de Modélisation Mathématique et Informatique de Systèmes COmplexes (UMMISCO)*, le *Laboratoire de Génétique des Microorganismes (LGM)* et au LIP6 l'équipe *Systèmes Multi-Agents (SMA)*. La présence de ces acteurs locaux est une opportunité fantastique pour le département, qui s'insère ainsi dans un tissu extrêmement riche avec lequel les complémentarités sont nombreuses et les interactions très prometteuses. D'un point de vue scientifique, le contexte est également riche en possibilités, puisque la thématique est en plein essor. Les attentes applicatives et sociétales sont extrêmement fortes, avec des domaines nouveaux comme la modélisation de la ville comme un système complexe, la gestion de la mobilité individuelle dans les réseaux, ou l'étude et l'exploitation des réseaux sociaux à grande échelle, par exemple. L'émergence de ces domaines est profondément liée à la possibilité de collecter, stocker et manipuler des données massives, problématiques qui sont au cœur de l'informatique et nous donnent donc un positionnement privilégié.



Vague D : Campagne d'évaluation 2012-2013

7.2 Projet 2014–2018 - Equipe APR

1 Membres permanent de l'équipe

Nom	Prénom	Status	Etablissement
Aubry	Philippe	MC	UPMC
Bui-Xuan	Binh-Minh	CR	CNRS
Chailloux	Emmanuel	PU	UPMC
Genitrini	Antoine	MC	UPMC
Pelletier	Maryse	MC	UPMC
Peschanski	Frédéric	MC	UPMC
Queinnec	Christian	PU	UPMC
Soria	Michèle	PU	UPMC
Trébuchet	Philippe	MC	UPMC
Valibouze	Annick	PU	UPMC

2 Evolution de l'équipe

Les fondations de l'équipe APR, créée en 2008, reposent sur trois piliers principaux : la diversité des parcours et des expériences de ses membres, la volonté de mettre à profit cette diversité dans le cadre de projets de recherche communs, et le lien étroit entre nos thématiques et projets de recherche et la formation universitaire. Après des débuts très encourageants en matière de publications et de projets financés, c'est avec une grande confiance dans le potentiel de nos travaux, ainsi qu'une très forte motivation que nous abordons l'avenir. La composition de l'équipe reflète un bon équilibre entre permanents et doctorants/post-doctorants. Cependant nous avons perdu une force vive avec la promotion de Olivier Bodini comme Professeur à l'université Paris 13, et deux de nos Maîtres de Conférences soutiendront leur HDR dans l'année 2012-2013. Il serait donc nécessaire de prévoir plusieurs recrutements pour assurer la pérennité de l'équipe.

3 Caractérisation de la recherche

Les thématiques de l'équipe concernent l'algorithmique et les langages de programmation, sous des aspects fondamentaux à la fois théoriques et pratiques, dans des domaines d'application qui vont de la programmation des nouvelles architectures de machines au développement de langages pour la concurrence et la mobilité, en passant par le test logiciel. Nous développons des approches méthodologiques et formelles, en sémantique (conception, implantation de langages, etc.) et en algorithmique (modélisation, analyse de complexité, etc.), pour parvenir jusqu'à des objectifs très concrets, dans la construction de prototypes logiciels.

Les interactions entre algorithmes et langages de programmation sont nombreuses et à différents niveaux : importation de concepts et de méthodes algorithmiques pour l'étude des langages, adaptation des schémas d'algorithmes aux langages utilisés ; et réciproquement, utilisation d'algorithmes de calcul complexes pour réaliser et tester le parallélisme ou la répartition dans les méthodes de programmation envisagées.

4 Projet scientifique

Les recherches de l'équipe s'organisent en quelques axes principaux autour de l'étude des structures discrètes, et de la conception et de l'implantation de langages et d'algorithmes permettant de construire des outils pour la réalisation d'applications sûres et efficaces. Nous articulons nos travaux en trois thèmes principaux: le thème 1 sur la génération aléatoire et l'analyse d'algorithmes ; le thème 2 sur la conception et l'implantation de langages pour les applications Web-critique ; le thème 3, sur le développement de langages et d'algorithmes pour le multicœur, est plus prospectif et doit renforcer encore nos interactions internes.

Positionnement par rapport aux thèmes transverses et de département Nos travaux s'intègrent naturellement dans plusieurs axes de prospectives transverses du LIP6, tout particulièrement en Optimisation (thème 1), Fiabilité des Logiciels (thème 2) et Nouveaux Systèmes Répartis (thème 3). La création prochaine du département "Système Complexes" est aussi un point très positif pour APR, et nous espérons multiplier les collaborations avec les équipes ComplexNetworks et MoVe.

4.1 Thème 1 : Génération Aléatoire et Analyse d'Algorithmes

Description Un objectif de la génération aléatoire est la simulation de grands modèles combinatoires. Il s'agit de générer des objets de très grande taille et en très grand nombre, pour visualiser des propriétés limites, modéliser des contextes réels ou faire des tests intensifs. Le développement récent de générateurs de complexité théorique linéaire, permet d'envisager un passage à l'échelle qui était impensable auparavant. La thématique de génération aléatoire est directement liée à l'analyse en moyenne des performances des algorithmes par l'étude des structures discrètes associées.

Originalité Nous développons dans APR des méthodes de Combinatoire Analytique [3] pour analyser les performances des algorithmes et faire de la génération aléatoire de structures complexes. Nos méthodes permettent de générer des objets combinatoires à partir de leur spécification, par des algorithmes génériques et efficaces. Nous travaillons à étendre l'expressivité de nos modèles, mais aussi à adapter les méthodes à des domaines applicatifs réels : recherche d'information dans les données massives, tests de robustesse ou validation de logiciels, modélisation de graphes, étude quantitative de la concurrence.

Positionnement dans le contexte local/ national/ international La communauté française qui travaille sur les structures aléatoires discrètes et l'analyse d'algorithmes est répartie dans plusieurs équipes fédérées dans le groupe ALEA du GDR-IM. De plus l'ANR MAGNUM est un ciment important pour les équipes de la région parisienne. Au niveau international, notre activité est bien reconnue, comme en atteste notre participation à plusieurs comités de sélection de revues et de conférences internationales. Nous avons par ailleurs des contacts bien établis avec plusieurs équipes qui travaillent sur les mêmes thématiques, avec lesquelles nous sommes en train de monter le réseau international *Alea-Research*.

Objectifs de résultats à 5 ans Nos objectifs concernent à la fois le développement de méthodes de combinatoire analytique, et les applications réelles, analyse d'algorithmes et génération aléatoire. Nous avons plusieurs projets ambitieux comme la génération non uniforme et à partir de spécifications multi-paramétrées, l'étude et la génération de graphes ainsi que l'analyse quantitative des langages de programmation. Tous ces travaux s'accompagneront d'implantations systématiques et efficaces dans des langages modernes.

4.2 Thème 2 : Conception et Implantation de Langages pour les applications Web-critique

Description Les langages déclaratifs ont pour objectif principal d'augmenter l'expressivité des programmes en garantissant la sûreté d'exécution sans perdre en efficacité. Ce tryptique expressivité/sûreté/efficacité représente un véritable défi pour les langages du futur, en particulier pour le développement Web, qui représente notre principal domaine d'application, ainsi que pour les outils de développement de logiciels critiques. Notre approche s'appuie sur l'utilisation des langages multi-paradigmes et nécessite d'adapter de nombreux traits de programmation à ces nouveaux domaines d'application.

Originalité Notre équipe se distingue par son positionnement à l'intersection entre les travaux théoriques (en collaboration avec le laboratoire PPS de l'Université de Paris Diderot, et avec l'équipes Indes de l'Inria-Sophia) et les aspects pratiques (développement de prototypes expérimentaux, transferts de technologie en partenariat avec des industriels). En guise d'illustration, nos travaux précurseurs [5] sur l'usage des continuations pour le développement d'applications Web ont inspiré de nombreuses technologies actuelles (cf. Apache Cocoon, Racket Web Server, etc.)

Positionnement dans le contexte local/ national/ international Nous collaborons étroitement avec deux projets, qui se caractérisent par leur envergure internationale : le projet Hop (Inria-Sophia), dans le cadre du projet ANR PWD (Programmation du Web Diffus), et le projet OCsigen (PPS, université Paris Diderot), dans le cadre initial d'une thèse en co-encadrement sur bourse Digiteo.

Nous collaborons aussi avec les industriels leaders mondiaux sur le développement de plates-formes logicielles, en particulier avec Adacore, Esterel-Technologies et Clearsy, dans le cadre du projet Couverture (FUI, Systematic) et du projet ANR CERCLES.

Objectifs de résultats à 5 ans L'enjeu pour les 5 ans à venir concerne particulièrement le développement de nos techniques - typage statique et vérification - permettant, dans le cadre des projets Hop et Ocsigen, d'augmenter la sûreté des applications Web de prochaine génération. Il nous faut aussi promouvoir la programmation déclarative pour le développement d'applications Web et de logiciels critiques, effort qui nécessitera l'appui de partenaires industriels, ainsi que le développement d'outils grand public.

4.3 Thème 3 : Algorithmes et Langages Déclaratifs pour le Multicœur

Description Un véritable besoin d'abstraction est né de la démocratisation du parallélisme et de la concurrence. L'objectif est de faciliter la composition parallèle - voire de l'automatiser - tout en permettant à un compilateur d'engendrer un code efficace pour une architecture cible donnée. Ces abstractions doivent aussi être spécifiées de façon claire et permettre l'analyse statique d'un certain nombre de propriétés pour garantir la sûreté d'exécution des programmes.

Originalité Ce thème - à la croisée de l'algorithmique (parallèle) et des langages déclaratifs - se distingue par son grand potentiel fédérateur. La grande force de notre équipe tient à sa complémentarité par rapport aux travaux que nous avons menés ces dernières années. Nous disposons d'un savoir-faire reconnu concernant les techniques avancées d'algorithmique et de programmation parallèle sur les architectures matérielles actuelles, notamment dans le modèle à mémoire partagée et le data-parallélisme. Nous étudions la théorie de la concurrence et de la mobilité à base de π -calcul, et par ailleurs nos travaux récents et novateurs concernant l'utilisation de la combinatoire analytique et la génération aléatoire pour l'analyse quantitative et la vérification probabiliste des langages parallèles, nous apparaissent comme très prometteurs. Enfin la partie combinatoire permet d'ajouter à l'étude de ces modèles de programmation, des analyses de performance et des simulations, et les applications de calcul fournissent un champ d'exemples complexes pour étudier la puissance et la robustesse de nos modèles.

Positionnement dans le contexte local/ national/ international Malgré les très nombreuses équipes travaillant sur la programmation pour les architectures multicœur, nos travaux sont compétitifs dans le domaine du déclaratif: notamment pour les extensions concurrentes de langages fonctionnels typés, et les travaux se trouvant à la lisière de la modélisation et la programmation. Au sein du LIP6, nous avons des liens étroits avec les équipes travaillant dans ce domaine (2 thèses en co-encadrement avec Pequan, 2 projets avec Regal et Alsoc/Pequan). Nous sommes en train de développer les collaborations internationales.

Objectifs de résultats à 5 ans Nous visons des résultats se fondant sur notre double compétence à la fois algorithmique et linguistique de la concurrence et du parallélisme. La finalisation de certains de nos algorithmes de calcul ainsi que de nos modèles de compilation ciblant les architectures multicœur homogènes et hétérogènes nous permettra de prendre du recul afin de proposer des abstractions de plus haut niveau. Nous avons aussi grand espoir, d'ici 5 ans, quant à l'aboutissement de nos techniques novatrices concernant la vérification probabiliste de programmes parallèles basée sur la génération aléatoire uniforme.

5 Analyse SWOT

Points forts : Forte cohésion de l'équipe, fondement de sa création, appuyée sur la variété des parcours et compétences de ses membres. Synergies avec plusieurs autres équipes du LIP6 : ALSOC, CN, MOVE, PEQUAN, REGAL. Travail fondamental conséquent, s'accompagnant de réalisations logicielles et d'outils de développement.

Points faibles : La diversité de l'équipe - qui fait sa force principale - peut également être un point faible. Il est important de continuer notre effort de cohésion scientifique et d'intégration des travaux de recherche. Par

ailleurs plusieurs MdC étant sur le point de soutenir leur HDR, et donc destinés à partir, il faut maintenir les compétences de l'équipe. Enfin plusieurs membres de l'équipe exercent, depuis des années, de lourdes charges d'administration pédagogique dans le LMD.

Risques liés au contexte : Les technologies du Web et du multicœur étant dictées par l'industrie, il existe un certain risque concernant leur adoption au sein d'un projet de recherche. Nos travaux théoriques récents visent des applications pratiques, mais des verrous théoriques importants doivent être levés.

Possibilités liées au contexte : Diversité de compétences permettant d'aborder des directions de recherche originales. Collaborations avec de nombreuses équipes académiques et acteurs industriels.

6 Mise en œuvre

Valorisation et recherche de moyens : Nous avons des partenariats bien établis avec des équipes de recherche et des industriels, par le biais de plusieurs projets académiques et collaboratifs et nous incitons nos jeunes maîtres de conférences à monter de nouveaux projets. Nous projetons aussi d'intensifier notre participation à des réseaux internationaux, pour faciliter les échanges et les demandes de financement international.

Diffusion des résultats (stratégie de publication) : Nous nous sommes efforcés de renforcer nos communications et publications dans les meilleures conférences et revues internationales, et cet effort doit être maintenu dans les années à venir. Les conférences nationales ne sont pas pour autant à négliger, car elles sont souvent une source de collaborations fructueuses.

Stratégie locale / nationale/ Internationale : Nous avons une participation très active dans l'animation scientifique (responsabilité au sein de GDR, organisation de séminaires récurrents et de conférences internationales) et prévoyons de continuer dans ce sens. Nous souhaitons aussi développer nos collaborations internationales actuelles (en particulier avec l'Europe (Barcelone, Munich, Vienne, Cracovie, Stockholm), l'Asie (Tokyo-Osaka, Taipei) et l'Amérique (New-York, Montevideo) et nous sommes aussi en train de monter un réseau international (ALEA-Research) autour de l'analyse des structures discrètes.

Implication future dans l'enseignement : L'implication de l'équipe dans l'enseignement a été très importante dans la période 2007-12. Nous continuons à nous investir sur les nouvelles maquettes du L et M, avec des créations (cursus par alternance en apprentissage dans la spécialité STL du master, propositions d'innovations pédagogiques en L et en M) et un renforcement des échanges internationaux au niveau du L (responsabilité du programme international) et du M (cursus et stages à l'étranger en STL).

7 Références

- [1] O. H. Cyrille Comar, Jérôme Guitton and T. Quinot. Formalization and comparison of mcdc and object branch coverage criteria. In *6th conference on Embedded Real-Time Software and Systems (ERTS)*, 2012.
- [2] P. W. Ezra Cooper, Sam Lindley and J. Yallop. Links: web programming without tiers. In *proceedings of FMCO 2006, LNCS 4709*, 2006.
- [3] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [4] R. Milner. *Communicating and mobile systems: the pi calculus*. Cambridge University Press, 1999.
- [5] C. Queinnec. Continuations and web servers. *Higher-Order and Symbolic Computation*, 17(4):277–295, 2004.



Vague D : Campagne d'évaluation 2012-2013

7.3 Projet 2014–2018 - Equipe ComplexNetwork

1 Membres permanents de l'équipe

Nom	Prénom	Statut	Établissement
Guillaume	Jean-Loup	MC	UPMC
Latapy	Matthieu	DR2	CNRS
Magnien	Clémence	CR1	CNRS
Tarissan	Fabien	MC	UPMC

2 Évolution de l'équipe

Bénédicte Le Grand a été promue professeur à l'université Paris 1 – Panthéon Sorbonne à l'automne 2013. Son remplacement rapide est important pour maintenir la taille de l'équipe. D'autre part, nous souhaitons favoriser le passage de la responsable de l'équipe au rang A (professeur ou DR) au LIP6.

3 Caractérisation de la recherche

Nous étudions une large famille d'objets interdisciplinaires, qui sont en relation étroite avec les applications : les graphes de terrain. L'originalité de notre approche consiste à embrasser les problématiques soulevées dans leur ensemble, dans une approche guidée par les objets. De façon complémentaire, nous ne nous refusons aucune méthode d'investigation, et établissons ainsi de nombreux liens entre disciplines (à l'intérieur de l'informatique et en dehors). Des allers-retours permanents entre questions transverses et étude de divers cas particuliers assurent la pertinence des questions fondamentales abordées et l'applicabilité de nos résultats.

4 Projet scientifique

Les travaux menés par l'équipe depuis sa création en 2008 ont fait émerger quelques problématiques particulièrement prometteuses et pour lesquelles nous sommes particulièrement bien positionnés : la détection de communautés, les phénomènes de diffusion, et la métrologie. Une question transverse est commune à toutes ces problématiques : celle de l'étude de la dynamique des graphes de terrain qui devient un axe nouveau et central de l'équipe. Nous avons pour ambition de nous imposer comme leaders à l'échelle internationale sur ces thèmes. Nos travaux s'intègrent naturellement dans le nouveau département *Systèmes Complexes*, en gestation, ainsi que dans plusieurs axes de prospective du laboratoire, soit par les objets étudiés (réseaux mobiles pour l'axe *Intelligence ambiante* ou réseaux sociaux pour l'axe *Social Computing*), soit par les problématiques soulevées (défis *Optimisation* et *Large Scale Data*).

4.1 Communautés et cœurs

Description. Exécuter la méthode de Louvain, qui n'est pas déterministe, un grand nombre de fois permet de calculer les *probabilités* que des paires de nœuds appartiennent à la même communauté. Regrouper les paires de nœuds qui ont une forte probabilité permet d'obtenir des *cœurs*. Ces cœurs, contrairement

aux communautés, sont très stables. Ils capturent une information très riche et ouvrent des perspectives (fondamentales et applicatives) très prometteuses, que nous allons explorer.

Originalité. L'étude des cœurs n'est envisageable que parce que la méthode de Louvain est très efficace et permet ainsi de s'abstraire du non-déterminisme inhérent aux méthodes heuristiques utilisées pour la maximisation de la modularité.

Positionnement dans le contexte local/ national/ international. L'équipe ayant contribué de façon majeure au développement de la méthode de Louvain, qui est à l'heure actuelle la meilleure disponible, nous sommes très bien positionnés sur ce domaine et sommes en contact avec tous ses principaux acteurs.

Objectifs de résultats à 5 ans. Nous visons à valoriser et exploiter pleinement la notion de cœurs, ainsi qu'à contribuer de façon majeure aux questions des communautés recouvrantes, dynamiques et/ou locales.

4.2 Phénomènes de diffusion

Description. Nous souhaitons fonder la théorie des phénomènes de diffusion sur l'étude de cas réels. Nous avons montré que les modèles classiques ne permettent pas de rendre compte des comportements observés. Nous proposerons donc des modèles plus pertinents basés sur nos observations. Par ailleurs, nous souhaitons étudier en profondeur l'impact de la dynamique du graphe sous-jacent sur les diffusions.

Originalité. L'équipe utilise une approche originale pour étudier les phénomènes de diffusion, fondée sur des captures d'échanges de fichiers dans des systèmes pair-à-pair. Ceci permet de suivre la diffusion simultanée d'un très grand nombre (plusieurs millions) de fichiers en même temps, ce qui est fondamental dans un contexte où l'acquisition de données est extrêmement ardue.

Positionnement dans le contexte local/ national/ international. Nous collaborons avec les chercheurs français utilisant une approche orientée données, notamment au travers du projet ANR WebFluence. Il existe au niveau international plusieurs autres groupes actifs étudiant les diffusions dans des contextes réels, notamment le groupe de Jure Leskovec à l'université de Stanford.

Objectifs de résultats à 5 ans. Nous visons à mettre au point une bonne caractérisation de phénomènes de diffusion réels, à évaluer de manière rigoureuse la pertinence des modèles existants, et à proposer des modèles plus pertinents.

4.3 Métrologie

Description. Nous proposons ici une toute nouvelle approche : mettre au point des processus de mesure ciblant une propriété donnée d'un graphe. En concevant une famille de tels processus, nous pouvons estimer de manière fiable plusieurs de ses propriétés. Allant plus loin, générer un graphe aléatoire ayant toutes ces propriétés permettra d'obtenir un graphe artificiel ressemblant plus au graphe réel que n'importe quelle carte. Par ailleurs, dans le cas de graphes dynamiques faire varier la fréquence et/ou la durée de mesure permet de mettre en évidence le biais induit par la dynamique sur les observations. Nous comptons ainsi tirer des conclusions rigoureuses sur la dynamique des graphes observés.

Originalité. L'approche usuelle consiste à étudier des *cartes* de graphes, qui donnent en fait une vision biaisée. L'approche orientée propriété est donc orthogonale par rapport à cette approche. La métrologie dynamique est quant à elle un sujet encore très peu abordé.

Positionnement dans le contexte local/ national/ international. L'équipe est déjà très bien positionnée sur le sujet. Nous avons entamé des collaborations avec les deux principales équipes françaises du domaine : l'équipe *Réseaux et protocoles* du LSIT, et l'équipe D-NET du LIP. Nous sommes en contact avec des équipes importantes au niveau international, comme le groupe de Mark Crovella à l'université de Boston.

Objectifs de résultats à 5 ans. Nous validerons l'approche orientée propriétés sur le cas particulier de la distribution des degrés de la topologie de l'internet, puis nous mettrons ensuite au point des méthodes plus génériques. Concernant la métrologie dynamique, nous visons à établir un ensemble de résultats rigoureux sur l'influence de la fréquence et de la durée de mesure.

4.4 Dynamique structurelle

Description. L'essentiel reste à faire concernant l'étude de la dynamique des graphes de terrain. Nous voulons proposer des méthodes permettant de la capturer de manière fiable, la décrire puis enfin la modéliser. Nous proposerons donc de nouvelles métriques permettant de caractériser la dynamique en adaptant les notions classiques mais aussi en concevant de nouvelles notions, puis étudierons les liens entre ces différentes métriques. Nous nous intéresserons également à l'échelle de temps à laquelle un graphe évolue. On peut en effet s'attendre à ce qu'il y en ait plus d'une dans le cas général, et nous avons montré que les observations varient selon l'échelle de temps à laquelle on étudie le graphe. Enfin, nous visons à proposer des modèles capables de reproduire les propriétés observées. Nos travaux préliminaires ont montré que des pistes à base de mélange des liens d'un graphe dynamique sont prometteuses.

Originalité. L'équipe est parmi les plus importantes au niveau mondial concernant la collecte de données sur la dynamique dans plusieurs contextes, et nous sommes pionniers dans leur étude. D'autre part, notre expertise sur plusieurs cas particuliers nous permet d'effectuer des études transverses.

Positionnement dans le contexte local/ national/ international. Il s'agit d'un sujet en pleine émergence au niveau international, et plusieurs équipes s'y intéressent. Cependant, aucun consensus n'existe encore pour la description et la modélisation des graphes dynamiques. L'équipe, grâce à ses travaux antérieurs, est idéalement située pour effectuer des contributions de grande portée.

Objectifs de résultats à 5 ans. Nous visons à proposer des notions générales pour l'étude de la dynamique, ainsi que des modèles pertinents de graphes dynamiques, basés sur l'étude de différents cas particuliers.

5 Analyse SWOT

Points forts. Nous nous situons dans un domaine émergent, qui gagne en reconnaissance au niveau international (présence dans les CFP des grandes conférences, apparition de revues et conférences dédiées). Nous sommes une des équipes les plus importantes sur ces thèmes en France et à l'international, et nos résultats antérieurs nous mettent en bonne position pour effectuer des contributions de premier ordre.

Points faibles. Notre approche vise à faire des allers-retours permanents entre les cas concrets et les questions fondamentales, nous couvrons donc un large spectre thématique. C'est une position difficile pour une équipe de petite taille. Le nombre de permanents de l'équipe doit rester raisonnable.

Risques liés au contexte. Nos sujets sont très actifs au niveau international ; il est donc difficile de rester compétitif.

Possibilités liées au contexte. Les contributions dans le domaine sont encore très dispersées. Il y a donc un fort besoin de contributions visant à unifier la science des graphes de terrain, ce qui est un de nos objectifs. Le contexte offre d'autre part de nombreuses opportunités de collaboration, tant dans différents domaines de l'informatique qu'avec des chercheurs d'autres disciplines.

6 Mise en œuvre

Valorisation et recherche de moyens. Nous avons déjà des collaborations avec des équipes académiques et des industriels qui ont été formalisées par des projets, et nous comptons continuer à en établir. Nous ferons aussi des demandes pour des financements particulièrement sélectifs, comme les *grants* de l'ERC.

Diffusion des résultats (stratégie de publication). Nous souhaitons continuer à publier dans les grandes conférences et revues de réseaux (Infocom, Transactions on Networking, ...), ainsi que dans les conférences et revues dédiées qui commencent à émerger. Nous avons par ailleurs pour ambition de publier nos résultats les plus marquants dans des revues à très haut impact comme *Science*, *Nature* et PNAS.

Stratégie locale / nationale/ internationale. Nous participons au Réseau National des Systèmes Complexes, à l'institut des systèmes complexes de Paris-Île de France, ainsi qu'au GDR ASR, au labex SMART, et au *flagship* FuturICT de l'union européenne. Nous comptons augmenter notre visibilité et l'impact de nos résultats par l'invitation de collègues prestigieux.

Implication future dans l'enseignement. Une nouvelle UE *Structure et Dynamique des Réseaux*, sous la responsabilité de J.-L. Guillaume, a été créée pour la nouvelle maquette du master. F. Tarissan sera co-responsable de la nouvelle première UE de licence, *Éléments de programmation*.



agence d'évaluation de la recherche
et de l'enseignement supérieur

Unités de recherche

Vague D : Campagne d'évaluation 2012-2013

7.4 Projet 2014–2018 - Equipe MOVE

1 Membres permanent de l'équipe

Nom	Prénom	Status	Etablissement
Baarir	Souheib	MC	UNIV PARIS 10
Bendraou	Reda	MC	UPMC
Besse	Cédric	MC	UPMC
Bérard	Béatrice	PU1	UPMC
Dutheillet	Claude	MC0	UPMC
Gervais	Marie-Pierre	PU1	UNIV PARIS 10
Hillah	Lom Messan	MC	UNIV PARIS 10
Ilié	Jean-Michel	MC0	IUT PARIS 5
Jaume	Mathieu	MC	UPMC
Kordon	Fabrice	PU1	UPMC
Legond-Aubry	Fabrice	MC	UNIV PARIS 10
Malenfant	Jacques	PU1	UPMC
Mounier	Isabelle	MC	UPMC
Paviot-Adet	Emmanuel	MC	IUT PARIS 5
Poitrenaud	Denis	MC	IUT PARIS 5
Pradat-Peyre	Jean-François	PU1	UNIV PARIS 10
Saint-James	Emmanuel	MC0	UPMC
Sznajder	Nathalie	MC	UPMC
Thierry-Mieg	Yann	MC	UPMC
Ziadi	Tewfik	MC	UPMC
Ziane	Mikal	MC	IUT PARIS 5

2 Evolution de l'équipe

La période 2007-2012 nous a permis de stabiliser, d'étendre et de pérenniser les résultats du quadriennal précédent. Notre taille est celle que nous avons lors de l'évaluation précédente : plusieurs recrutements de qualité (en majorité des permanents extérieurs à l'UPMC) sont venus compenser des promotions et des départs en retraite. Cette taille autorise un travail d'exploration en profondeur et permet d'obtenir des résultats significatifs, par le biais de coopérations portant sur des aspects théoriques et des réalisations pratiques.

La taille de l'équipe étant celle d'un petit laboratoire, la question d'un éventuel découpage émerge. Si la vision "naturelle" consiste à distinguer les activités de vérification des activités de modélisation, c'est cependant une solution de facilité qui nuirait à un aspect original que nous présentons en section suivante. Ainsi, lors de la période qui s'ouvre, une réflexion s'engagera sur la pertinence de maintenir MoVe dans sa configuration actuelle ou l'essaimer dans de nouvelles équipes.

3 Caractérisation de la recherche

Nos travaux de recherche sont à la fois de nature fondamentale, avec un travail théorique sur les modèles formels et les questions d'expressivité et de décidabilité de la vérification qui s'y rapportent, et également de nature appliquée, avec la définition de langages de modélisation et la réalisation d'outils efficaces de modélisation et de vérification utilisés pour résoudre des cas d'études. Nous obtenons ainsi une "boucle vertueuse" : *i*) nous traitons des études de cas et constatons les limites de nos modèles ou de nos outils, *ii*) nous réfléchissons à de nouvelles approches de modélisation ou de vérification, *iii*) nous développons une nouvelle génération d'outils plus performants ou mieux adaptés à ces situations nouvelles.

4 Projet scientifique

L'émergence planifiée du département Systèmes Complexes que nous rejoindrons nous place au cœur d'une thématique en pleine expansion: l'analyse de systèmes complexes. Ces derniers sont répartis (c'est le thème pour lequel nos approches sont particulièrement efficaces) mais doivent aussi respecter d'autres contraintes comme des échéances temporelles ou la robustesse.

À ce titre, les activités de l'équipe MoVe et leurs récents développements, se situent au cœur d'une activité de fiabilisation des logiciels complexes:

- le besoin en outils et approches de vérification formelle efficaces (i.e. capables de passer à l'échelle s'accroît, ce qui est la préoccupation du thème méthodes et outils pour le *model checking* efficace),
- l'exploration de solutions afin de savoir si elles satisfont certains critères, l'analyse de leur robustesse ou de leur fonctionnement en mode dégradé sont les préoccupations principales du thème vérification quantitative, contrôle et synthèse,
- enfin, l'insertion de toutes ces techniques formelles dans un processus de fabrication, leur association avec des notations dédiées à un domaine d'application (i.e. confortables pour les spécialistes du domaine) ainsi que le suivi des procédures de développement sont au centre des travaux du thème ingénierie logicielle pilotée par les modèles.

Nos activités se rattachent à deux thèmes transversaux du LIP6: *i*) sécurité et fiabilité du logiciel (nos travaux d'appliquent dans les deux aspects couverts par cette thématique) et *ii*) nouveaux systèmes répartis (nos techniques de modélisation et d'analyse sont particulièrement optimales dans ce cas précis).

Au niveau national, nos activités nous impliquent dans le GDR GPL (Génie des Procédés Logiciels) et dans l'action inter-GDR AFSEC (Approches Formelles des Systèmes Embarqués Communicants).

4.1 Thème 1: méthodes et outils pour le *model checking* efficace

Description. Pour résister au passage à l'échelle, les outils doivent faire cohabiter différentes approches de vérification. Pour cela, ils doivent pouvoir passer d'un algorithme à un autre en fonction de la spécification du système et de ses propriétés. Combattre l'explosion combinatoire requiert aussi une approche modulaire de la vérification couplée avec des techniques d'abstraction.

Pour connecter les outils de vérification à des langages de plus haut niveau (points d'entrée pour la vérification comme FIACRE, AADL, etc. ou certains profil UML) il faut disposer d'un "assembleur pour la vérification" à la fois neutre vis-à-vis du langage d'origine et véhiculant les informations nécessaires à nos algorithmes les plus efficaces.

Originalité. Nous faisons partie des quelques équipes, au niveau national et international, à avoir exploré une multitude d'approches de vérification comme les diagrammes de décision, l'usage des symétries ou l'utilisation de différents types d'automates. Nous avons en particulier combiné certaines approches dans des prototypes aux performances très encourageantes. C'est là un travail à la fois théorique et pratique (aspects développement importants).

La participation régulière au *model checking contest* a démontré les performances de nos approches sur des spécifications de systèmes parallèles face aux outils concurrents.

Positionnement dans le contexte local/ national/ international. Nous nous positionnons en pionniers dans l'encodage hiérarchique de diagrammes de décision, que nous pratiquons depuis maintenant un peu plus de cinq ans. Nous maîtrisons cette technique et nos prototypes deviennent de plus en plus matures.

Notre expertise est également reconnue dans la conception et la réalisation d'outils de vérification. Nos outils et bibliothèques sont utilisés par la communauté et cités en référence [5, 3, 4].

Nos collaborations en région parisienne se renforceront, notamment grâce à la plate-forme CosyVerif, lancée en 2012 avec le LSV et le LIPN. Nous comptons intensifier les collaborations avec l'IRCCyN et le LAAS sur l'analyse temporisée dans le cas discret (en relation avec le thème 2).

Au niveau international, nos collaborations avec l'Université de Turin sur l'analyse stochastique et avec l'Université de Genève sur l'exploitation des types de données abstraits seront amenées à s'intensifier. Notre participation régulière au *model checking contest* ouvre des possibilités de collaborations avec les équipes de recherche intéressées par le développement d'outils.

Objectifs de résultats à 5 ans. Nous identifions les objectifs suivants:

- passer de la maîtrise de différentes approches (et leur combinaison lorsque c'est possible) à une forme plus intégrée du *model checking* adaptatif. SPOT est la base d'un *framework* le permettant
- comparer nos techniques aux approches SAT/SMT pour le *model checking*. L'objectif est de nous comparer pour des spécifications ayant les mêmes caractéristiques mais aussi de transposer certaines stratégies entre les deux approches.
- sur la base de notre expérience (*Instantiable Transition Systems*), développer un nouvel "assembleur pour la vérification" gérant mieux les aspects relatifs aux données et à la dynamique des systèmes.
- examiner l'application potentielle de nos démarches au temps dense (*e.g.* exploitation des symétries et leur couplage avec des diagrammes de décision).

4.2 Thème 2: vérification quantitative, contrôle et synthèse

Description. Les systèmes complexes doivent non seulement être fiables mais aussi robustes [2]. En effet, plongés dans des environnements de plus en plus divers, leur conception doit leur permettre de faire face à des situations imprévues lors de leur utilisation, de manière à tolérer de petites perturbations. Le contrôle et la synthèse de systèmes robustes est un enjeu majeur et l'analyse quantitative peut produire des mesures de la robustesse ou la caractérisations des paramètres. Les modèles temporisés et/ou probabilistes sont donc des cibles privilégiées pour ce travail.

Originalité. Nous abordons les questions de robustesse sous deux angles originaux : *i*) dans le contexte du projet ANR ImpRo, nous produisons un cadre théorique unifié dans lequel exprimer les problèmes de robustesse et explorer leurs liens avec les contraintes d'implémentation et *ii*) suite au projet ANR FME3, nous approfondissons les aspects de la robustesse relatifs aux circuits soumis à des perturbations électromagnétiques.

Positionnement dans le contexte local/ national/ international. Les travaux prévus seront menés dans le premier cas avec des chercheurs de l'IRCCyN et du LSV et dans le second cas avec l'équipe SOC du LIP6 et avec le laboratoire TIMA. Une coopération avec l'Ecole Polytechnique de Montréal (projet CFQCU) est également prévue sur les aspects quantitatifs de propriétés de sécurité.

Objectifs de résultats à 5 ans. Nous proposons :

- l'extension des résultats obtenus pour la synthèse de systèmes asynchrones, en introduisant éventuellement des contraintes de résistance aux pannes ;
- l'approfondissement de l'étude du contrôle et de la synthèse pour les modèles temporisés, en intégrant les objectifs de robustesse ;
- la définition de mesures de robustesse, en lien avec les questions d'opacité et de diagnostic ;
- l'addition d'informations stochastiques sur des diagrammes de décision, pour y étendre les méthodes d'agrégation markovienne et les calculs d'indices de performance (en relation avec le thème 1) ;
- l'élaboration de techniques d'abstraction pour l'analyse de robustesse de circuits de taille significative.

4.3 Thème 3: Ingénierie logicielle pilotée par les modèles

Description. Récemment standardisée en 2008, la modélisation multi-vues permet de simplifier la conception d'un système complexe en le scindant en plusieurs vues. Des difficultés empêchent cependant de tirer pleinement profit de cette approche. Citons par exemple la simulation par l'expert métier du système modélisé, lui permettant de vérifier des propriétés. Cela pose le problème de l'alignement sémantique

des langages dédiés utilisés pour les différentes vues et de leur prise en main par des experts métiers (non informaticiens).

Originalité. Avec l'apparition des langages dédiés, l'expert métier devient un acteur du développement des logiciels complexes. Toutefois, son rôle est encore principalement cantonné à la modélisation. Nous visons à élargir son action jusqu'à la vérification de propriétés par simulation du modèle du système. En considérant l'aspect multi-vues, cela nécessite de développer des environnements dédiés supportant l'exécution et le raisonnement sur des modèles hétérogènes.

Positionnement dans le contexte local/ national/ international. Cette préoccupation de formaliser les langages dédiés à des fins de vérification par simulation directement par les experts métiers émerge depuis peu dans la communauté. Mais les travaux restent focalisés sur chacune des parties du problème, sans intégration de solutions dans un environnement à disposition des experts métiers. Par exemple, l'unification sémantique de modèles hétérogènes a fait l'objet de l'action spécifique GeMoC du GDR GPL (Génie de la Programmation et du Logiciel) [1]. La définition de langages dédiés est traitée dans les universités de Mannheim et de Wellington et également. Les aspects notation graphique sont également étudiés par LIFL et au LIG. Nous sommes en relation avec ces différents acteurs, notamment à travers notre participation à l'action spécifique "End-User Modelling" du GDR GPL.

Objectifs de résultats à 5 ans. Nous visons à fournir aux experts du domaine un environnement support à la production logicielle et assurant le suivi méthodologique des approches multi-vues. Cela inclut :

- L'alignement sémantique des différents langages utilisés pour modéliser les vues.
- la génération du code global du système à partir de plusieurs vues et donc de plusieurs générateurs associés.
- Assurer le suivi méthodologique (i.e., procédé de réalisation) dans le cas d'une approche multi-vues.
- Mettre les outils de modélisation entre les mains d'experts du domaine de chaque vue, qui sont souvent des non-informaticiens.

5 Analyse SWOT

Points forts : pionniers sur les diagrammes de décision hiérarchiques, expertises sur de nombreuses techniques pour combattre l'explosion combinatoire du *model checking*, reconnaissance internationale de nos outils.

Points faibles : le sujet est jugé important mais est très compétitif (présence d'excellentes équipes en région parisienne).

Risques liés au contexte : l'approche outil requise par notre "cercle vertueux" est difficile à tenir (besoin de stabilité de nos bibliothèques et outils), en particulier dans un contexte économique particulièrement frileux chez les industriels. La mutualisation de l'effort sur plusieurs laboratoires (projet CosyVerif) est pour nous une réponse à ce risque.

Possibilités liées au contexte : fort potentiel de collaboration via le groupe MeFoSyLoMa¹, impacts de nos résultats potentiellement important, besoins croissant dans les domaines où nos travaux sont applicables.

6 Mise en œuvre

Valorisation et recherche de moyens : la mutualisation de la valorisation d'outils grâce à la plate-forme Cosyverif devrait nous ouvrir des opportunités de partenariat avec les industriels. Nous avons demandé en juin 2012 l'intégration du projet au sein de l'IRILL (demande en cours).

Diffusion des résultats (stratégie de publication) : nous avons déjà accru la diffusion de nos résultats dans des revues de renommée internationales. Nous continuerons mais sans négliger les conférences majeures de notre domaine comme TACAS, ATVA, CAV, CONCUR, ICSE, MODELS, etc.

Stratégie locale / nationale/ Internationale : notre implication dans deux thèmes transverses du LIP6 nous permet de côtoyer les labex associés à ces termes. Nous envisageons également des collaborations avec le Labex DigiWorlds (Saclay, architectures et données distribuées). Nos relations avec l'IRCCyN (ANR

1. Méthodes Formelles pour les Systèmes Logiciels et Matériels, groupant les équipes de vérification de 8 laboratoires en région Parisienne, <http://www.mefosyloma.fr>

IMpro) et le LAAS seront amenées à s'intensifier, en particulier via de nouveaux projets ANR. Notre participation à l'organisation du *model checking contest* devrait accroître notre visibilité.

Implication future dans l'enseignement : MoVe restera adossée à la spécialité SAR (Systèmes et Applications Répartis) du master d'informatique. Nous avons aussi des relations étroites avec la spécialité STL (Science et technologie du Logiciel). Nous sommes aussi très actifs dans la gestion et la réforme du L avec plusieurs UE centrales (Outils mathématiques pour l'informatique, UML, etc.) animée par les membres de MoVe.

7 Références

- [1] B. Combemale, C. Hardebolle, C. Jacquet, F. Boulanger, and B. Baudry. Bridging the Chasm between Executable Metamodeling and Models of Computation. In *5th International Conference on Software Language Engineering (SLE)*, Lecture Notes in Computer Science, page to appear. Springer, 2012.
- [2] T. A. Henzinger. Two challenges in embedded systems design: Predictability and robustness. *Philosophical Transactions of the Royal Society A*, 366:3727–3736, 2008.
- [3] S. Hostettler, A. Marechal, A. Linard, M. Risoldi, and D. Buchs. High-level petri net model checking with alpina. *Fundamenta Informaticae*, 113(3-4):229–264, 2011.
- [4] J. I. Requeno and J. M. Colom. Compact representation of biological sequences using set decision diagrams. volume 154 of *Advances in Intelligent and Soft Computing*, pages 231–239. Springer, 2012.
- [5] D. Tabakov and M. Y. Vardi. Optimized temporal monitors for systemc. In *First International Conference on Runtime Verification (RV)*, volume 6418 of *Lecture Notes in Computer Science*, pages 436–451. Springer, 2010.

Chapitre 8

Projet scientifique - Département SOC

Sommaire

8.1	Projet 2014 – 2018 - Département SOC	142
8.2	Projet 2014–2018 - Equipe ALSOC	144
8.3	Projet 2014–2018 - Equipe CIAN	149
8.4	Projet 2014–2018 - Equipe SYEL	154



Vague D : Campagne d'évaluation 2012-2013

8.1 Projet 2014 – 2018 - Département SOC

1 Présentation du département

Le département concentre l'ensemble de ses recherches sur les méthodes et outils de conception des systèmes intégrés sur puce. Il rassemble une centaine de personnes et est structuré en 3 équipes de recherche : les travaux de l'équipe ALSOC portent sur le développement de méthodes et outils pour la conception et l'utilisation d'architecture de processeurs multi-cores. L'équipe CIAN s'intéresse au développement de méthodes de conception des composants numériques et analogiques pouvant être intégrés dans des systèmes sur puce (AMS SOC). Les recherches de l'équipe SYEL visent à la conception de systèmes électroniques et de leurs applications.

Les applications visées par le département s'inscrivent dans le domaine de l'électronique et informatique embarqués (transports, santé, télécom) et l'informatique grand public. Le département entretient des coopérations avec de nombreux partenaires industriels (ST Micro-electronics, Thalès, Bull, Orange Labs, NXP, Philips..) et académiques nationaux (INRIA, Institut Télécom, LIRMM, LSV, SUPELEC, TIMA, UTC..) et internationaux (Ain Shams Université, Universitat Bremen, Wien TU..).

2 Périmètre scientifique

Les recherches menées dans le département sont principalement à vocation applicative, ce qui explique l'importance des contrats de recherche et des liens très étroits avec le milieu industriel. Sa richesse réside dans la diversité des compétences de ses membres et des travaux qui y sont menés : de l'électronique à l'informatique, des aspects circuits aux problèmes liés à leur utilisation, en passant par les aspects modélisation et simulation de ces systèmes complexes.

Le département est un acteur majeur du GdR SOC/SIP sur les thématiques architectures (manycores et reconfigurables), électronique (AMS et systèmes hétérogènes), sécurité, vérification et consommation. Il entretient également des liens avec les GdR ISIS, ASR, GPL et RO.

3 Positionnement

- Implication dans les thèmes transverses *Nouveaux Systèmes Répartis, Informatique Ambiante/Objets Communicants, Fiabilité et e-santé*.
- Participation au Labex SMART dans le programme « Autonomie humaine et e-santé ».
- Implication dans les activités du GdR SOC/SIP : Patrick Garda est directeur du GdR. Franck Wajsburt est co-responsable du thème « Logiciel embarqués et Architectures Matérielles » et est membre du comité de pilotage. Le colloque 2012 a été organisé par le département sur le site de Jussieu.
- Principaux laboratoires académiques concurrents ou partenaires en France : LIRMM, TIMA, Lab-STICC, LTCI, INL, . . .
- Principaux laboratoires académiques concurrents ou partenaires à l'étranger : Fraunhofer IIS/EAS, EECS (Berkeley), AGRA (Bremen), EPFL (Lausanne), Katholieke Universitat (Leuven), Technische Universiteit (Eindhoven), Politecnico di Torino, Karlsruhe Institute of Technology, Rice University, Ecole Polytechnique de Montréal . . .

4 Les défis scientifiques

Les avancées technologiques ont permis l'émergence de systèmes sur puce avec des densités d'intégration dépassant couramment le milliard de transistors et des finesses de grilles nanométriques. Le développement d'outils pour la conception de systèmes intégrés sur puce fiables et sûrs et leur utilisation nécessite de relever des défis majeurs qui fédèrent les membres du département.

Systèmes hétérogènes Les technologies de fabrication permettent aujourd'hui d'intégrer sur une même puce des systèmes comportant des traitements numériques analogiques et des micro-capteurs/actionneurs multi-physiques. Un premier défi est de développer un flot de conception capable de supporter toutes les parties de la puce et leurs communications. Le département peut le relever grâce à son expertise dans le développement d'outils pour la conception analogique, numérique, MEMS, et dans les méthodes de simulation.

Fiabilité et sécurité des circuits et systèmes L'évaluation de la robustesse d'un circuit ou système vis à vis d'un critère donné (vieillesse, variabilité, fautes transitoires/permanentes, politiques de sécurité), ainsi que l'établissement de contre-mesures adaptées, deviennent incontournables du fait de la miniaturisation des composants et de leur utilisation dans des contextes sécuritaires. Le département peut proposer des solutions intéressantes en combinant ses compétences sur les architectures robustes, les outils de synthèse, les méthodes formelles et l'analyse dynamique de code.

Architectures manycore Le développement de flots de conception pour une architecture manycore composée de plusieurs centaines de processeurs et d'outils qui tirent parti du nombre de processeurs élevé est un des grands défis de l'informatique. Le département peut y apporter des réponses pertinentes grâce au développement et l'utilisation d'outils de conception, et de son expertise dans le développement d'OS embarqué, d'architectures, de réseaux sur puce et de méthodes d'optimisation.

Architectures reconfigurables Ces architectures ont connu un essor considérable en termes de complexité et de performances. Le développement sur une même puce de systèmes FPGA et de processeurs pour les reconfigurer dynamiquement pose des nouveaux défis liés à leur conception et leur utilisation. L'expertise du département doit permettre le développement de solutions pertinentes optimisant leur consommation et l'utilisation des ressources disponibles.

Technologies émergentes L'organique, la bionique, la photonique, les technologies mémoires telles que les MRAM ou la 3D posent de nouveaux défis pour les réaliser et les concevoir. Le département réunit les compétences nécessaires pour développer des méthodes, architectures et outils pour la modélisation, la conception, et la validation des systèmes basés sur ces technologies émergentes et tenant compte des aspects liés à leur consommation, leur fiabilité et leur sécurité.

Systèmes embarqués pour la santé Ces systèmes regroupent un ensemble de dispositifs électroniques qui peuvent être implantés dans un milieu biologique sans le dégrader. Le grand défi aujourd'hui est de mettre en place des outils méthodologiques pour les concevoir, ce qui requiert une collaboration étroite entre des spécialistes de la santé et des systèmes embarqués. Il est accessible aux chercheurs du département qui ont investigué ce domaine, et crée des liens avec le secteur médical.

5 Evaluation du département : analyse SWOT

Points forts Compétences complémentaires des équipes sur la plupart des aspects liés à la conception et la réalisation de systèmes sur puce, combinant des recherches applicatives et des aspects plus théoriques. Doubles compétences en électronique et en informatique. Bonne visibilité grâce aux collaborations entre les équipes, la participation à plusieurs projets nationaux et internationaux d'envergure, et la diffusion de plateformes de référence. Implication importante dans la diffusion des connaissances.

Points faibles Dynamique de publication et présence dans les comités de programme à renforcer. Insuffisance du personnel d'encadrement (Prof./DR), de chercheurs à plein temps et d'ingénieurs.

Risques liés au contexte Forte dépendance des décisions stratégiques des industriels du domaine. Forte compétition régionale (Labex DigiWorlds, IRT SystemX), nationale et internationale.

Possibilités liées au contexte Domaine en pleine expansion avec des enjeux économiques et sociétaux conséquents. Implication dans le Labex SMART. Attractivité liée à la localisation et à l'UPMC.



Vague D : Campagne d'évaluation 2012-2013

8.2 Projet 2014–2018 - Equipe ALSOC

1 Membres permanents de l'équipe

Nom	Prénom	Status	Etablissement
Bouyer	Manuel	IR	UPMC
Braunstein	Cécile	MC	UPMC
Desbarbieux	Jean-Lou	MC	UPMC
Drach-Temam	Nathalie	PR	UPMC
Encrenaz	Emmanuelle	MC HDR	UPMC
Genius	Daniela	MC	UPMC
Greiner	Alain	PR	UPMC
Heydemann	Karine	MC	UPMC
Marchetti	Olivier	MC	UPMC
Meunier	Quentin	MC	UPMC
Munier	Alix	PR	UPMC
Pecheux	François	MC HDR	UPMC
Wajsbürt	Franck	MC	UPMC

2 Evolution de l'équipe

Le départ en retraite de A. Greiner aura lieu vers 2015-2016.

3 Caractérisation de la recherche

La recherche effectuée dans l'équipe est majoritairement à visée applicative ; les outils formels et algorithmes développés sont expérimentés sur des applications réelles dans le cadre de projets institutionnels.

4 Projet scientifique

Le "[développement d'] architectures performantes intégrant totalement les aspects hardware et software pour accroître fonctionnalités, disponibilité et fiabilité." est un axe prioritaire mentionné dans le dernier rapport de Stratégie Nationale pour la Recherche et l'Innovation (2009) [3].

Au travers de son offre d'outils de modélisation et conception de systèmes many-cores intégrés sur puce, l'équipe ALSOC (et plus généralement le département SOC) s'inscrit pleinement dans cet axe ; l'équipe est un partenaire reconnu dans la communauté scientifique et industrielle de conception de systèmes intégrés sur puce. Pour la période 2014-2018, l'équipe ALSOC devrait rester sur son coeur de compétences (conception matériel/logiciel des architectures et systèmes many-cores embarqués sur puce). En s'appuyant sur les développements déjà réalisés dans la période 2007-2012, l'équipe va intégrer de nouvelles dimensions dans les outils qu'elle propose et évaluer ses derniers dans des contextes industriels au travers des projets collaboratifs auxquels l'équipe est associée. Les collaborations intra-département (SOC) devraient être renforcées sur les thématiques suivantes : modélisation de systèmes multi-physique, analyse de robustesse de circuits/systèmes.

Les collaborations transverses aux autres départements du LIP6 se retrouvent essentiellement dans les thèmes *Nouveaux Systèmes Répartis, Fiabilité, et Informatique Ambiante/Objets Communicants*.

4.1 Thème 1 : Modélisation et simulation de systèmes complexes

Description Les systèmes embarqués et architectures many-core interagissent avec un environnement physique analogique (communication RF, capteurs biologiques...). La modélisation précise de ces interactions est un sujet de recherche d'actualité, nécessitant de repenser le couplage des domaines discrets (monde numérique) et continu (monde analogique) [2]. Par ailleurs, les conditions environnementales influent sur le comportement du système et doivent être appréhendées lors du prototypage, par exemple l'influence de la température sur le vieillissement des composants et leur dérive de fonctionnement. En outre, la complexité des systèmes à analyser requiert de très grandes puissances de calcul et la simulation parallèle est un recours [9]. Par ailleurs, les techniques de simulation actuellement en oeuvre seront améliorées en vue d'une parallélisation du coeur du simulateur. L'équipe ALSOC poursuivra le développement de la plateforme de prototypage virtuel SOCLib pour intégrer ses différents domaines ; elle participera aux recherches de l'axe transversal "Informatique Ambiante/Objets Communicant" pour la modélisation d'environnement complexes à bse de réseaux de capteurs.

Originalité Les techniques mises en oeuvre visent à l'intégration dans un même environnement de descriptions de domaines différents, en s'appuyant sur l'extension du langage SystemC au monde analogique – SystemC-AMS.

Positionnement La plate-forme de modélisation SOCLib est utilisée par une centaine d'utilisateurs dans le monde, dont STMicroelectronics qui souhaite l'utiliser pour la conception de sa machine multiprocesseur P2012 (projet BGLE/ARGILE). Par ailleurs, l'équipe ALSOC a contribué à la standardisation du langage de modélisation SystemC-AMS qui permet d'adjoindre des descriptions de systèmes analogiques (variables continués régies par des équations différentielles ordinaires) aux systèmes numériques. Concernant la modélisation de systèmes hétérogènes, les équipes ALSOC et CIAN ont développé des collaborations avec Fraunhofer IIS/EAS Dresde (Allemagne), TU Vienna (Autriche), TU Kaiserslautern (Allemagne), TU Delft (Pays-Bas), EPFL (Suisse), TIMA (France), INL (France), Université Cantabria (Espagne), InESS Strasbourg (France), IEMN (France) ; les projets collaboratifs FP7-VERDI et Catrene H-INCEPTION soutiendront ces développements. Les principaux compétiteurs sont Univ. Southampton (T. Kazmierski, Grande-Bretagne), Univ. Berkeley (E. Lee, USA), KTH (A. Jantsch, Suède).

Objectifs de résultats à 5 ans L'équipe souhaite disposer d'une plate forme de modélisation et simulation de systèmes hétérogènes intégrant différents domaines physiques.

4.2 Thème 2 : Définition et conception d'architectures many-core à mémoire partagée

Description Les architectures many-core à mémoire logiquement partagées nécessitent la mise en oeuvre d'un mécanisme de cohérence, qui peut être assuré par le système d'exploitation [8] ou par des caches matériels cohérents [7] ; l'approche proposée par l'équipe ALSOC au travers du projet TSAR est une cohérence assurée par le matériel, exécutant un système d'exploitation dont les structures de données sont distribuées ; ce type de structure *réduisant la contention* est particulièrement adapté pour la programmation d'applications s'exécutant sur ces architectures. De nombreux travaux concurrents sur les systèmes d'exploitation adaptés aux many-core sont effectués aux États-Unis (projet Corey (MIT) [5]) et en Europe (projet BARELFISH (ETH) [4]). Actuellement, un démonstrateur de la machine TSAR est disponible (4 clusters) ; la définition et implantation des premiers services du système d'exploitation est opérationnelle (système ALMOS). Dans les 5 ans à venir, cette activité sera poursuivie et plusieurs fonctionnalités seront ajoutées à l'architecture TSAR et au système d'exploitation ALMOS : cohérence mémoire (protocole moins consommateur d'énergie et plus rapide), système d'exploitation distribué avec système de fichiers, I/O distribuées, tolérance aux pannes franches au travers d'un mécanisme de détection et reconfiguration.

Originalité : l'architecture TSAR se distingue des propositions concurrentes (de Intel, Tilera ou Kalray) sur le maintien de la cohérence en matériel et l'utilisation d'un protocole write-through scalable : cette proposition offre le modèle de programmation le plus simple (et compatible avec la majorité des codes PC existants), avec un protocole moins complexe que les stratégies de type write-back. Les expérimentations

sur le premier prototype ont montré la faisabilité de l'approche mais de nombreuses questions restent irrésolues, notamment la réalisation d'un système de fichier distribué efficace, la consommation du système et la stratégie de recouvrement en cas d'éléments défaillants, qui seront investiguées dans les 5 ans à venir.

Positionnement Ces travaux sont supportés par le projet Medea+ SHARP ; des études comparatives par rapport à l'architecture concurrente P2012 pourront être menées suite au projet BGLE/ARGILE (dans lequel l'architecture P2012 de STM-CEA/LETI sera modélisée avec SOCLib). La comparaison avec les solutions de maintien de la cohérence de façon logicielle (sur le 48-coeurs d'Intel) pourront être évaluées dans le projet transversal du LIP6 "Nouveaux Systèmes Répartis".

Objectifs de résultats à 5 ans L'équipe souhaite réaliser un prototype à 128 coeurs de l'architecture TSAR intégrant les mécanismes de sécurité et tolérance aux pannes, implanté sur multi-FPGA, exécutant le système d'exploitation ALMOS disposant d'un système de fichiers distribué et une gestion scalable des entrées/sorties. Les cartes FPGA seront entièrement développées par BULL. En cas d'échec de BULL, la démonstration sera faite par la synthèse de 24 coeurs sur 2 FPGAs (déjà disponibles au LIP6).

4.3 Thème 3 : Déploiement et adaptativité d'applications sur architectures many-cores

Description La programmation d'applications tirant partie de la puissance de calcul et du passage à l'échelle qu'offrent les architectures many-core reste un problème difficile. Se posent aussi les problèmes de contention (accès aux ressources partagées) et de dimensionnement optimal des ressources (buffer mémoire). Le modèle de programmation data-flow se généralise au travers de la notion de composants [10], et l'usage de mémoires transactionnelles offre des services de synchronisation de haut niveau [6] ouvrant des pistes pour aborder le problème. Au demeurant, le besoin de stratégies de placement des tâches et données mémoires adaptatives reste pressant. De telles stratégies permettraient par exemple de prendre en compte les interactions entre applications ou de s'adapter aux données imprévisibles statiquement. L'équipe propose deux approches pour résoudre ces problèmes. La première concerne le développement de stratégies adaptatives de parallélisation (division de tâches ou non / nombre de tâches selon le modèle de programmation/de parallélisme), de placement et d'ordonnancement des tâches parallèles et/ou de préchargement de données associées (pour architecture NUMA) en se basant sur la collecte d'informations dynamiques (compteurs matériels, sondes logicielles, observateurs) ; certaines stratégies pourront être intégrées dans le système ALMOS (th2). La seconde consiste à développer de nouveaux modèles permettant de prendre en compte les applications non régulières. Les problèmes d'optimisation sont alors modélisés sous forme de programme linéaires en nombres entiers et étudiés avec des outils de la recherche opérationnelle.

Enfin la sécurité des applications doit être garantie vis-à-vis d'attaques matérielles ou de fautes transitoires. L'analyse dynamique de l'application peut être utilisée pour détecter des comportements particuliers (identifiés préalablement comme des attaques potentielles ou des défaillances, à caractériser [1]) et activer des réactions pré-établies à ces comportements. L'identification de ces divergences de comportement repose sur la modélisation de l'impact des fautes matérielles au niveau du programme en langage d'assemblage ou du circuit au niveau RTL, et la formalisation de chemins d'attaques qui exploitent les relations entre les grandeurs observables du programme assembleur en cours d'exécution avec les informations protégées.

Originalité Le projet se base sur l'analyse de corrélations entre évolution des compteurs matériels et les comportements significatifs de l'application et d'autre part sur la mise en oeuvre d'approches analytiques pour la résolution de problèmes d'optimisation complexes. Concernant la sécurité, les fautes matérielles sont considérées et on s'intéresse à l'interaction matériel/logiciel et à la mise en oeuvre de contre-mesures matérielles, en se basant sur des modèles formels pour stratégies d'activation des contre-mesures.

Positionnement Ces thématiques sont étudiées en collaboration avec d'autres équipes du laboratoire au travers de projets LIP6, ainsi qu'avec les équipes TRACES-IRIT, INRIA-Parkas, IBISC, CEA-LSC sur ces différents aspects.

Objectifs de résultats à 5 ans L'équipe souhaite proposer des techniques de parallélisation adaptative en fonction des ressources et des comportements des accès mémoire ; intégrer les algorithmes d'optimisation

dans le compilateur SigmaC de KalRay et proposer des algorithmes efficaces pour applications à comportement irrégulier ; proposer, en collaboration avec le CEA, les bases d'une plateforme de monitoring de code, détection de déviations (pré-établies) et réactions online dans le contexte des attaques matérielles.

5 Analyse SWOT

Points forts : L'équipe bénéficie de compétences transverses combinant de la recherche applicative et des aspects plus théoriques, avec le soucis des applications, pour mener à bien ces thèmes de recherche ; SOCLib et TSAR disposent d'une bonne visibilité ainsi que du support de plusieurs projets nationaux et internationaux. Les trois thèmes s'inscrivent dans les projets transverses du laboratoire.

Points faibles : Le départ en retraite d'A. Greiner vers 2015-2016 risque d'affaiblir la dynamique du thème 2. Le thème 3 regroupe des activités de recherche plus fondamentale et n'est pas actuellement supporté par un projet national ou international. Dans les trois thèmes, la dynamique de publication pourrait être renforcée (actuellement, le ratio "revues et conférences de rang A / revues et conférences internationales" est d'environ 1/3).

Risques liés au contexte : L'aspect *intégration sur puce* nous rend assez dépendants des décisions stratégiques des acteurs français et européens du domaine (Bull, STMicroelectronics, Thales). Au niveau régional, le Labex DigiWorlds implanté à Saclay autour des systèmes embarqués couvre les thématiques du projet.

Possibilités liées au contexte : Le contexte économique reste favorable : la conception de systèmes embarqués est prégnante dans la société et les aspects de certification vont croissant, nécessitant des méthodes de conception de pointe ; plusieurs appels à projets nationaux et internationaux couvrent nos axes de recherche (ANR, Catrene, FP7). Plusieurs laboratoires sur Paris recoupent les thématiques du projet (CNAM, LIENS, INRIA/AOSTE). Au niveau de l'UPMC, le Labex SMART a des thématiques connexes aux thèmes 1 et 3 du projet.

6 Mise en œuvre

Valorisation et recherche de moyens : La valorisation des outils se poursuivra au travers de logiciels libres protégés par des licences de type LGPL. Nous poursuivrons la politique de collaboration avec académiques et industriels au travers de projets, en cherchant à diversifier nos partenaires industriels.

Diffusion des résultats (stratégie de publication) : On intensifiera l'effort de publication et visera prioritairement la publication des résultats dans les conférences de référence et journaux internationaux du domaine (revues ACM-TODAES, ACM-TCAD, ..., conférences DAC, DATE, HiPEAC, ICCAD, ...); la liste des conférences du domaine ainsi que leur pertinence a été élaborée conjointement avec les équipes CIAN et SYEL, et servira de référentiel pour la soumission des articles à publier.

Stratégie locale / nationale/ Internationale : Une implication forte dans le thème prospectif transverse "Nouveaux Systèmes répartis" permet de confronter les choix de l'équipe à d'autres solutions de maintien de la cohérence (thème 2), ce qui pourrait encore être renforcé au travers de collaborations avec les participants du Labex DigiWorlds (Saclay : architectures et données distribuées). Un développement de partenariat au sein de l'UPMC (Labex SMART autour des systèmes embarqués et robotique) ou avec le laboratoire LIENS (ENS, équipe Parkas) serait profitable aux thèmes 1 et 3. En particulier, les collaborations émergentes avec INRIA-AOSTE, IBISC et LTCI-LabSOC seront renforcées.

Implication future dans l'enseignement : Nous avons contribué à l'élaboration de la nouvelle maquette de Licence et Master d'informatique, et dans le parcours SESI (Systèmes Electroniques Systèmes Informatiques) de ce master, nous avons défini un coeur de compétences autour des "architectures multi-coeurs, OS et applications", à visée professionnalisante ou académique. L'établissement d'échanges internationaux (comme au travers de projets PHC-Procope) sera encouragée.

7 Bibliographie

- [1] J. Balasch, B. Gierlichs, and I. Verbauwhede. An in-depth and black-box characterization of the effects of clock glitches on 8-bit mcus. In *8th International Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC)*, IEEE, 2011.

- [2] A. Benveniste, T. Bourke, B. Caillaud, and M. Pouzet. Non-standard semantics of hybrid systems modelers. *J. Comput. Syst. Sci*, 78(3), 2012.
- [3] M. de l'Enseignement Supérieur et de la Recherche. Stratégie nationale de recherche et d'innovation, 2009. http://media.enseignementsup-recherche.gouv.fr/file/SNRI/69/8/Rapport_general_de_la_SNRI_-_version_finale_65698.pdf.
- [4] A. B. et al. The multikernel: a new os architecture for scalable multicore systems. In *ACM SIGOPS 22nd symposium on Operating systems principles*, 2009.
- [5] S. B.-W. et al. Corey: An operating system for many cores. In *8th USENIX Symposium on Operating Systems Design and Implementation*, 2008.
- [6] V. Gramoli, R. Guerraoui, and V. Trigonakis. Tm2c: A software transactional memory for many-cores. In *EuroSys*, 2012.
- [7] K. Martin, M. Dill, and D. Sorin. Why on chip cache coherence is here to stay. *Communication of the ACM*, 55(7), 2012.
- [8] J. Park, C. Jang, and J. Lee. A software-managed coherent memory architecture for manycore. In *2011 International Conference on Parallel Architectures and Compilation Techniques (PACT)*, 2011.
- [9] J. Peeters, N. Ventroux, T. Sassolas, and L. Lacassagne. A SystemC-TLM framework for distributed simulation of complex systems with unpredictable communication. In *DASIP*, 2011.
- [10] A. Pop and A. Cohen. Expressiveness and data-flow compilation of openmp streaming programs. In *INRIA Technical Report*, 2012.



Vague D : Campagne d'évaluation 2012-2013

8.3 Projet 2014–2018 - Equipe CIAN

1 Membres permanent de l'équipe

Nom	Prénom	Status	Etablissement
Aboushady	Hassan	MC	UPMC
Anceau	François	PEM ¹	CNAM
Bazargan Sabet	Pirouz	MC	UPMC
Chaput	Jean-Paul	IE	UPMC
Chotin-Avot	Roselyne	MC	UPMC
Galayko	Dimitri	MC	UPMC
Iskander	Ramy	MC	UPMC
Louërat	Marie-Minerve	CR1	CNRS
Mehrez	Habib	PU2	UPMC
Porte	Jacky	CHB ²	TélécomParisTech
Renault	Patricia	MC ³	UPMC

2 Evolution de l'équipe

CIAN accueille actuellement un nombre important de doctorants et souhaiterait recruter un Professeur pour renforcer sa capacité d'encadrement ainsi que sa capacité de prise de responsabilité (équipe, département SOC, master SESI, spécialité ELelectronique-Informatique (ELI) de Polytech'Paris-UPMC). Plusieurs enseignants-chercheurs soutiendront leur HDR dans l'année 2012-2013. Il serait donc nécessaire de prévoir plusieurs recrutements pour assurer la pérennité de l'équipe.

3 Caractérisation de la recherche

Les travaux de CIAN portent sur l'étude des phénomènes physiques qui impactent les circuits analogiques, numériques et mixtes pour en déduire des principes de fonctionnement fondamentaux et construire des applications complexes. L'équipe CIAN souhaite garder un bon équilibre entre les travaux théoriques (principes physiques, développement d'algorithmes et méthodes de conception) et les preuves de concepts sous la forme de réalisations logicielles et matérielles (FPGA, fabrication de circuits intégrés CMOS, nouvelles technologies).

4 Projet scientifique

L'équipe CIAN, forte de sa double compétence en informatique et en micro-électronique, a acquis une solide expérience dans le domaine des architectures, des méthodes et des outils pour la modélisation, la simulation, la conception et la vérifications des circuits analogiques, numériques et mixtes.

1. professeur émérite
2. chercheur bénévole
3. en disponibilité

L'équipe continuera à proposer des solutions aux nouveaux défis engendrés par les deux axes que sont la miniaturisation accrue (« More Moore ») et l'hétérogénéité de plus en plus affirmée (« More than Moore »). Ces dernières années, de nouvelles technologies ont émergé (3D, nouveaux dispositifs micro-électro-mécanique, optiques, organiques). Elles vont constituer des atouts majeurs pour résoudre de nombreux problèmes mais vont apporter également des défis auxquels il faut apporter des solutions. On notera, en particulier, la progression croissante de la technologie 3D qui présente une excellente opportunité pour résoudre certains problèmes nanométriques mais également un réel bouleversement pour l'intégration de circuits hétérogènes.

Les collaborations transverses aux autres départements du LIP6 se retrouvent essentiellement dans les thèmes *Fiabilité/Sécurité*, et *Informatique Ambiante/Objets Communicants*.

4.1 Architectures et CMOS « More Moore »

Description. *Sécurité* : Les concepteurs d'algorithmes de cryptographie ont besoin de savoir rapidement si l'implantation matérielle de leur algorithme sera performante. En effet, cela devient un critère primordial pour les organismes de normalisation tels que le NIST (National Institute of Standards and Technology), comme on a pu le voir pour le standard de chiffrement (AES) et aujourd'hui pour celui de la fonction de hachage (SHA3). Il devient donc important de faciliter le prototypage rapide vers des cibles matérielles pour les concepteurs d'algorithmes de cryptographie au moyen d'un langage de description qui leur soit proche et familier, contrairement au projet Athena [2] de l'université américaine George Mason qui propose un environnement de synthèse vers FPGA d'algorithmes de cryptographie décrits en VHDL. Forte de son expérience en synthèse de chemins de données arithmétiques, l'équipe pourra relever ce défi. Également lié à la sécurité, nous voyons émerger une nouvelle technologie : l'informatique quantique. La conception de circuits quantiques pour la cryptographie est un défi sur lequel nous avons commencé à travailler et que nous souhaitons approfondir. Il s'agit ici d'adapter les méthodes de conception, que l'équipe maîtrise parfaitement, en prenant en compte les principes de la mécanique quantique (superposition d'états, intrication etc.).

Fiabilité : Avec les technologies actuelles, nous voyons de plus en plus apparaître des problèmes liés à la fiabilité des circuits. Comme l'annonçait S. Borkar[1] en 2004, les circuits revenant de fonderie vont comporter de plus en plus de défauts et d'autres défauts apparaîtront au cours du fonctionnement. Il devient plus qu'urgent de tenir compte de ces problèmes dès la phase de conception de ces circuits. Jusqu'ici, la phase de synthèse ne tenait compte que de la surface, du temps de propagation et de la consommation pour optimiser une architecture. L'équipe propose d'ajouter ce nouveau critère de fiabilité aux algorithmes d'optimisation qu'elle a développés et d'étudier l'impact de la fiabilité sur les autres critères.

D'autres défis sont à relever et concernent, en particulier, les problèmes d'électro-migration, le vieillissement et la variabilité. Des architectures, des méthodes et des outils doivent être développés, pour que, de façon précoce et dès la conception, les circuits puissent être analysés puis rendus plus robustes à ces phénomènes.

La fiabilité des circuits dépend aussi de leurs mécanismes de communication interne : nous poursuivons les recherches sur la génération d'horloge, adaptée aux circuits CMOS en technologies très avancées. Le développement de l'approche *Globally Synchronous Locally Synchronous* (GSLs) permet de contourner les problèmes liés à la métastabilité. Toutefois, l'évolution de la technologie laisse entrevoir l'émergence d'autres problèmes de fiabilité comme l'occurrence de perturbations géantes dans les circuits d'alimentation. Celles-ci risquent de provoquer des « pannes sans défaut » susceptibles de perturber gravement le comportement de ces circuits. Leur étude devrait constituer un passage obligé à la réalisation de circuits complexes de haute fiabilité.

Radio cognitive : La radio cognitive est basée sur des composants portables capables d'analyser l'encombrement du spectre et d'établir une communication sans fil sur les bandes de fréquences disponibles avec les standards adéquats. Actuellement, la consommation et la surface des circuits capables d'établir le type de communication sont loin d'être adaptées à des appareils portables. Notre équipe a une grande expérience dans la conception et la réalisation de convertisseurs analogiques-numériques (CAN) Sigma-Delta passe-bande RF. Les résultats de mesure de 4 circuits déjà fabriqués montrent que ces CAN ont une consommation 10 fois plus faible que l'état de l'art dans le domaine [7].

Architectures Reconfigurables : Les architectures reconfigurables connaissent et continueront à avoir un essor considérable [5]. Le dernier circuit de chez Xilinx, le Virtex 7, comporte environ 7 milliards de transistors avec une technologie avancée proche du 3D et cette tendance va s'affirmer dans les années à venir. L'équipe a une grande expérience dans l'optimisation des architectures et des outils pour les circuits FPGA. Elle n'a pas la prétention de rivaliser avec les grands fabricants de ces circuits mais elle vise à contribuer à des solutions qui permettront de réduire le fossé entre les circuits ASIC et FPGA. Elle vise en particulier l'intégration d'unités reconfigurables dans les SoCs pouvant être utilisées comme coprocesseurs. Certaines « niches » sont encore à creuser telles que la tolérance aux pannes et la robustesse de ces unités, la sécurité face aux attaques, des architectures avec des topologies de réseaux d'interconnexion optimisés, l'exploitation du 3D avec des outils de partitionnement performants etc.

Originalité. Adaptation du flot de conception à d'autres contextes (cryptographie, fiabilité, quantique). Nouvelles applications très prometteuses dans le domaine des « smart phones », tablettes mais aussi pour la sécurité automobile et les réseaux de capteurs sans fil.

Positionnement dans le contexte local/ national/ international. Collaborations avec les concepteurs d'algorithmes de cryptographie (CalSci au LIP6), des équipes reconnues pour leurs travaux sur la fiabilité (TIMA, Telecom ParisTech).

Nouvelle architecture pour le front-end RF déjà bien avancée avec des circuits fabriqués et mesurés. Collaboration avec d'autres équipes du LIP6 pour la réalisation des couches PHY et MAC dédiées à la radio cognitive.

Objectifs de résultats à 5 ans. Nous proposerons un environnement de conception dont le langage de description sera proche des concepteurs d'algorithmes cryptographiques et qui leur permettra de synthétiser efficacement leurs algorithmes vers de l'ASIC ou du FPGA. Nous proposons également d'introduire la fiabilité dans les critères d'optimisation au même titre que les critères habituels.

Nos perspectives en Radio Cognitive sont :

- de réaliser un récepteur RF complet basé sur ces Sigma-Delta RF avec les circuits numériques de translation de fréquence et de démodulation.
- de collaborer avec d'autres équipes (PHARE et NPA) du département réseaux du LIP6 afin de réaliser les couches PHY et MAC d'un récepteur RF basse-consommation dédié à la radio cognitive.

4.2 Technologies émergentes et « More than Moore »

Description. *3D* : Dans les prochaines années, les technologies 3D vont avoir une croissance très importante et vont permettre de résoudre certains problèmes nanométriques tels que la puissance dissipée et les performances des systèmes intégrés [4]. Elles vont rendre également possible une réelle intégration de circuits hétérogènes (numériques, analogiques, capteurs et actionneurs MEMS, RF, optiques, photovoltaïques etc.). Ces technologies représenteront un véritable bouleversement dans les méthodologies de modélisation, de conception et de validation. Certains défis concerneront notamment les problèmes de fiabilité, de distribution des alimentations et de variation de phase sur les horloges, de l'évacuation thermique etc.

Energie et consommation sur puce : La gestion de la consommation et la génération d'énergie dans les systèmes autonomes sont de principaux défis pour les concepteurs pour la décennie à venir. Les thèmes de gestion d'énergie sur puce, les interfaces avec des sources d'énergie et systèmes de conditionnement de convertisseurs d'énergie garderont leur actualité. L'équipe CIAN envisage donc de poursuivre les travaux dans cet axe riche en découvertes que ce soit sur le plan de la théorie des systèmes ou vis-à-vis des applications. Nous nous intéresserons également aux circuits de récupération d'énergie de l'environnement à l'échelle microscopique.

Organique : Les semiconducteurs organiques ont permis le développement d'une variété d'appareils électroniques faible coût sur substrats souples en plastique, y compris des cellules solaires, des écrans à diodes électroluminescentes, des transistors à effet de champ, des capteurs, des actionneurs et des transistors pour mémoires non volatiles. D'autre part, les dispositifs organiques souffrent d'un fonctionnement à basse fréquence, de grandes variations des paramètres du procédé de fabrication ainsi que d'une faible fiabilité et d'un mauvais rendement. Il est donc nécessaire de développer de nouvelles structures de circuits adaptées aux caractéristiques de cette technologie [6].

Outils et méthodes pour les architectures hétérogènes : Les systèmes sur puce (SoC) et les systèmes multi-puce (SiP) intègrent des parties numériques (matérielles et logicielles) mais aussi des parties analogiques/Radio Fréquence ainsi que des micro-capteurs/actionneurs (MEMS) multi-physiques. Ce sont des systèmes hétérogènes que l'on appelle « Cyber Physical Systems » (CPS). Ils intègrent sur un même substrat une application complexe. Il n'existe pas de flot de conception allant des spécifications « au niveau système », jusqu'au dessin des masques pour un CPS incluant numérique (matériel et logiciel), analogique, RF et MEMs [3].

Originalité. Equipe bénéficiant des compétences en physique, électronique et informatique pour étudier l'impact des nouvelles technologies sur les méthodes de conception et les applications.

Positionnement dans le contexte local/ national/ international. Forte collaboration entre le LIP6-CIAN en conception de circuits intégrés et l'école Polytechnique LPICM dans la fabrication des transistors organiques. A un niveau d'abstraction élevé, l'extension AMS du langage SystemC développée par l'OSCI a déjà donné de bons résultats. L'équipe CIAN va poursuivre la coopération avec le groupe de travail d'ASI (Accellera System Initiative, fusion des consortiums Accellera et OSCI) « AMS Working Group ».

Objectifs de résultats à 5 ans. Le 3D, nécessitant des compétences en architectures et en développements logiciels, représentera un axe fédérateur pour l'équipe voire transversal avec d'autres équipes. En effet, le 3D permettra de combler les limites du 2D pour des architectures complexes telles que les manycores. Dans ce sens, des coopérations pourront se faire avec l'équipe ALSOC. Au niveau des coopérations extérieures, d'ores et déjà, un programme de recherches sur 4 ans a été établi avec le CEA-DAM pour développer un FPGA 3D optimisé et sécurisé. D'autre part, la poussée des architectures hétérogènes, en combinaison avec notre savoir faire en algorithmique du partitionnement, offrira l'opportunité à l'équipe de fédérer les compétences des différents groupes numériques, analogiques, RF et capteurs. L'objectif sera de réaliser des systèmes 3D dont les empilements seront réalisés en bonne « intelligence » en fonction de la fonctionnalité de chaque partie et en tenant compte des problèmes thermiques.

Nous poursuivons l'étude des circuits intégrés haute-tension ultra-basse-consommation permettant de gérer l'énergie dans des capteurs/systèmes embarqués biomédicaux, microcapteur sans fil, etc. Nous étudierons la fiabilité des zones basse-tension intégrées avec des zones haute-tension dans le cadre du projet européen FP7 « AUTOMICS ».

Nous allons profiter de l'expérience de notre équipe dans la conception de circuits CMOS pour développer de nouvelles topologies et architectures capables de réaliser des circuits performants et fiables utilisant l'électronique organique. Ce projet est en collaboration avec le laboratoire LPICM de l'école Polytechnique.

Au niveau système, l'équipe CIAN va proposer des extensions CPS à SystemC, au niveau proche de la simulation électrique, et continuera à formaliser un processus de propagation des spécifications qui permette de raffiner les modèles d'un niveau d'abstraction élevé à un niveau d'abstraction plus bas, pour aboutir au dimensionnement des blocs analogiques (environnement CHAMS). Ce travail est soutenu par le projet européen Catrene « H-INCEPTION » en cours de financement, en coopération avec l'équipe ALSOC.

5 Analyse SWOT

Points forts : La double compétence informatique et électronique de l'équipe. Collaborations déjà établies avec des équipes françaises de pointe. Expériences accumulées par l'équipe (chaîne de CAO VLSI Alliance, environnement CORIOLIS).

Points faibles : Pour des raisons de coût de fabrication nous utilisons une technologie mature (CMOS 130nm) qui limite les performances.

Risques liés au contexte : Forte compétition nationale et internationale (France : IMS, LETI, Europe : IMEC Twente, Etats Unis : Broadcom, Intel, Japon : Tokyo Univ). Difficulté pour participer au pôle de compétitivité « Systematic » situé sur le plateau de Saclay, pour les thématiques CIAN.

Possibilités liées au contexte : Collaborations déjà établies avec des équipes internationales de pointe. Collaboration fructueuse avec ST sur des CAN à très hautes performances dans des technologies récentes. Nouvelles applications pour de l'électronique grande surface, faible coût et à substrat flexible. Facilité pour l'équipe CIAN de coopération avec les autres départements du LIP6 du campus Jussieu.

6 Mise en œuvre

Valorisation et recherche de moyens : Partenariat avec fondateurs européens (STMicroelectronics, AustriaMicroSystems, Centre Multi Projet). Projets européens débutant ou en cours d'acceptation : FP7 AUTOMICS (dont CIAN est le coordinateur) et Catrene H-Inception (STMicroelectronics est le coordinateur).

Diffusion des résultats (stratégie de publication) : Conférences en CAO : DATE et DAC. Participation aux TPC, proposition de tutoriels.

Stratégie locale / nationale/ Internationale : Participation au GDR SoC SiP. Membre du consortium de standardisation « Accellera System Initiative ». Renouer les relations avec les acteurs importants comme le pôle Systematic.

Implication future dans l'enseignement : Participation aux discussions sur la nouvelle maquette de licence d'Informatique et licence d'Electronique à l'UPMC. Participation aux discussions sur la nouvelle maquette du Master d'Informatique, spécialité SESI « Systèmes Electroniques, Systèmes Informatiques ». Implication dans les enseignements de la spécialité EElectronique-Informatique (ELI) de Polytech'Paris-UPMC. Membre du programme européen « Academic Program » pour l'enseignement de SystemC-AMS⁴.

7 Références

- [1] S. Borkar. Microarchitecture and Design Challenges for Gigascale Integration. In *Proceedings of the 37th annual IEEE/ACM International Symposium on Microarchitecture (MICRO-37)*, 2004.
- [2] K. Gaj, J.-P. Kaps, V. Amirineni, M. Rogawski, E. Homsirikamol, and B. Brewster. ATHENA - Automated Tool for Hardware Evaluation: Toward Fair and Comprehensive Benchmarking of Cryptographic Hardware using FPGAs. In *International Conference on Field Programmable Logic and Applications (FPL)*, pages 414–421, Aug. 2010.
- [3] H. Graeb. ITRS 2011 Analog EDA Challenges and Approaches. In *Design Automation and Test in Europe Conference (DATE)*, pages 1150–1155, march 2012.
- [4] S. K. Lim. Physical Design for 3D System on Package. *IEEE Des. Test*, 22(6):532–539, Nov. 2005.
- [5] J. Luu, I. Kuon, P. Jamieson, T. Campbell, A. Ye, W. M. Fang, K. Kent, and J. Rose. VPR 5.0: FPGA CAD and architecture exploration tools with single-driver routing, heterogeneity and process scaling. *ACM Trans. Reconfigurable Technol. Syst.*, 4(4):32:1–32:23, Dec. 2011.
- [6] H. Marien, M. Steyaert, E. van Veenendaal, and P. Heremans. Analog Building Blocks for Organic Smart Sensor Systems in Organic Thin-Film Transistor Technology on Flexible Plastic Foil. *J. Solid-State Circuits*, 47(7):1712–1720, 2012.
- [7] E. Martens, A. Bourdoux, A. Couvreur, R. Fasthuber, P. V. Wesemael, G. V. der Plas, J. Craninckx, and J. Ryckaert. RF-to-Baseband Digitization in 40 nm CMOS With RF Bandpass Delta-Sigma Modulator and Polyphase Decimation Filter. *J. Solid-State Circuits*, pages 990–1002, 2012.

4. <http://www.systemc-ams.org/educational.html>



Vague D : Campagne d'évaluation 2012-2013

8.4 Projet 2014–2018 - Equipe SYEL

1 Membres permanents de l'équipe au 1 septembre 2012

Nom	Prénom	Statut	Etablissement
ALEXANDRE	Annick	MCU	UPMC
DENOULET	Julien	MCU	UPMC
FERUGLIO	Sylvain	MCU	UPMC
GARDA	Patrick	PU1	UPMC
GRANADO	Bertrand	PU2	UPMC
HACHICHA	Khalil	MCU	UPMC
PINNA	Andrea	MCU	UPMC
VALLETTE	Farouk	MCU	UPMC

2 Evolution de l'équipe

Deux évolutions structurelles importantes vont se produire pour l'équipe SYEL. D'une part, sa composition change en septembre 2012 : Bertrand Granado rejoint l'équipe comme Professeur UPMC et Sylvain Viateur, technicien, quitte l'équipe pour une promotion comme ingénieur d'études. D'autre part, l'équipe SYEL devrait déménager vers la barre 24-25 en 2014, ce qui la rapprochera des autres équipes du LIP6.

3 Caractérisation de la recherche

L'équipe se positionne à la fois sur des aspects fondamentaux et appliqués. D'une part, elle contribue aux avancées fondamentales dans le domaine des systèmes électroniques, principalement leur modélisation et leur architecture. D'autre part, elle exploite ces avancées pour la conception de systèmes dédiés à une classe d'applications, principalement dans le domaine de la e-santé, à des fins de validation et de valorisation. Ces applications nourrissent en retour ses recherches fondamentales de nouvelles questions.

4 Projet scientifique

- *Evolutions par rapport à la période précédente* L'équipe SYEL poursuivra son positionnement sur les systèmes électroniques en les abordant d'un point de vue système et en allant vers les systèmes embarqués en réseaux, prenant en compte l'évolution du domaine. Sur le plan fondamental, les recherches se poursuivront vers la *modélisation des performances des systèmes embarqués*, spécifiquement l'estimation de la consommation et la modélisation à haut niveau de l'intégrité du signal. Par ailleurs, un nouveau thème de recherche sur les *architectures reconfigurables* sera développé par Bertrand Granado en synergie avec les autres thèmes de l'équipe. Sur le plan appliqué, l'équipe arrêtera ses activités dans le domaine du radar logiciel et de la radio-logicielle à la fin du projet SurfOnHertz, le départ d'Olivier Romain en 2011 la privant d'expertise dans le domaine des télécommunications. Par ailleurs, l'équipe développera ses activités dans le domaine de la santé et les fera évoluer du biomédical vers la *e-santé*. L'apprentissage automatique et les réseaux de capteurs sans fil feront le lien entre ces trois thèmes.

– *Positionnement par rapport aux thèmes transverses et de département* Les thèmes fondamentaux de l'équipe contribuent évidemment aux thèmes transverses du département SOC. Plus précisément, la modélisation des performances des systèmes embarqués contribue au thème transverse *systèmes hétérogènes* du département, en particulier l'estimation de la consommation. Les travaux sur l'intégrité du signal contribuent aux thèmes transverses *systèmes hétérogènes* et *fiabilité et sécurité*, abordée ici par la robustesse des circuits analogiques. D'un autre côté, le thème des architectures reconfigurables contribue au thème transverse sur les *architectures reconfigurables*.

Le thème e-santé de l'équipe SYEL contribue naturellement au thème transverse *e-santé* du LIP6. Des actions transversales avec les équipes ACASA (projet WiPo) et MALIRE (projet CareDiem) du département DAPA ont d'ailleurs commencé en 2012 grâce au soutien des projets LIP6. De plus, les travaux de l'équipe SYEL sur les réseaux de capteurs sans fil, abordés par l'estimation de la consommation et la conception de réseaux de capteurs biomédicaux, la conduisent naturellement à participer au thème transverse du LIP6 sur *l'intelligence ambiante et les objets communicants*.

4.1 Modélisation des systèmes hétérogènes

Description L'équipe SYEL poursuivra ses travaux sur la modélisation des performances des systèmes hétérogènes. Elle a pris contact avec l'équipe MALIRE pour améliorer la construction de ses modèles par apprentissage automatique. En ce qui concerne l'estimation de la consommation, l'évolution sera de développer un environnement pour estimer la consommation d'un réseau de systèmes embarqués, incluant des sources d'énergie. Les projets ANR WiNoCoD (avec NXP et l'équipe ALSOC) et CORAIL AME (avec THALES Avionics dans le cadre des Initiatives d'Excellence) apporteront des applications pour cet environnement. En ce qui concerne l'intégrité du signal, des contacts ont été pris avec l'équipe de Sonia Ben Dhia au LAAS dans la perspective d'applications dans le domaine avionique. Par ailleurs, les travaux sur la robustesse des fonctions analogiques pourraient se poursuivre vers une validation expérimentale (avec l'équipe CIAN du LIP6).

Originalité L'originalité tient au positionnement sur une méthode générique de modélisation à haut niveau d'abstraction des fonctions numériques, analogiques et radio. En effet, les travaux concurrents sont focalisés soit sur les fonctions numériques, soit sur des circuits analogiques à plus bas niveau d'abstraction.

Positionnement dans le contexte local / national / international Pour l'estimation de la consommation, les principaux concurrents sont l'IRISA, le Lab-STICC, le LEAT et l'INL sur le plan national, et KUL sur le plan international : l'équipe SYEL a une certaine avance sur l'estimation de la consommation des fonctions analogiques et des systèmes hétérogènes. Pour l'intégrité du signal, les concurrents sont le LAAS et l'Institut Polytechnique de Turin : l'équipe SYEL se situe à plus haut niveau d'abstraction.

Objectifs de résultats à 5 ans Ces objectifs sont de construire un environnement, qui permette d'estimer la consommation et l'intégrité du signal d'un réseau de systèmes embarqués incluant des sources d'énergie très tôt dans le processus de conception, et de valider expérimentalement la robustesse d'un oscillateur.

4.2 Architectures reconfigurables

Description L'équipe SYEL reprendra ses travaux sur les architectures reconfigurables et l'adéquation algorithmes-architecture. Les recherches porteront sur la définition d'architectures reconfigurables dynamiquement adaptées au support d'un système d'exploitation, l'architecture OLLAF notamment, et sur leur exploitation au sein d'un système distribué massivement parallèle. La mise en oeuvre matérielle de fonctions de communication distribuée, inspirées de MPI, sera étudiée afin de contribuer à la définition d'un modèle de calcul adapté à ces architectures. Des mécanismes bio-inspirés seront investigués pour réaliser des services de système d'exploitation pour ordonnancer les tâches de traitement, qu'elles soient logicielles ou matérielles, ou pour équilibrer la charge des ressources de calcul. Ces recherches seront menées en collaboration avec l'Université de Douala, l'ENSTA de Brest et l'Université Libre de Bruxelles.

Originalité Un des aspects novateurs de ces recherches est de comprendre comment rendre efficace un système d'exploitation sur une architecture dynamiquement reconfigurable, en prenant comme point de vue la conception d'une architecture spécifiquement adaptée. De plus, à l'inverse d'autres travaux de recherche, le positionnement est clairement celui d'un système distribué massivement parallèle où les ressources reconfigurables sont des ressources directement monopolisables et non de simples accélérateurs matériels.

Positionnement dans le contexte local / national / international L'équipe participe à l'animation du GT « Architectures Reconfigurables » du GDR SOC/SIP et elle construit des collaborations avec l'Université Libre de Bruxelles.

Objectifs de résultats à 5 ans Les objectifs à 5 ans sont de concevoir une architecture bio-inspirée auto-reconfigurable dynamiquement à l'aide d'un système d'exploitation matériel.

4.3 e-santé

Description L'équipe va amplifier ses travaux dans plusieurs directions. Tout d'abord, l'étude de capteurs implantés se poursuivra avec des capteurs implantés pour l'analyse de la moelle épinière, en relation avec le LIF. Ensuite, le projet CYCLOPE se poursuivra vers la conception de capsules pour l'analyse de maladies intestinales (avec Lariboisière). Par ailleurs, la compression d'images pour la e-santé se développera vers les images de signaux médicaux comme les EEG (projet FUI STIVE avec CIRA et Lariboisière). De plus, de nouvelles pistes émergent de collaborations internes au LIP6 avec les équipes ACASA (projet WiPo sur la polysomnographie sans fil avec Tenon) et MALIRE (projet Carpe Diem sur les capteurs pour l'informatique affective).

Originalité L'originalité de l'équipe réside dans l'appui de ses recherches en e-santé sur ses compétences dans les systèmes électroniques et ses travaux sur l'estimation de la consommation et les architectures reconfigurables. En effet, les équipes actives dans la e-santé sont le plus souvent soit des équipes de conception de circuits intégrés, auxquelles il manque une vision système, soit des équipes de Génie Biomédical, auxquelles il manque la maîtrise des méthodes de conception de systèmes.

Positionnement dans le contexte local / national / international L'équipe participe activement à l'animation de la communauté académique, localement au sein des axes transverses e-santé du LIP6 et du Labex SMART, et nationalement au sein du groupe Bionique commun aux GDR ISIS, SOC/SIP et STIC-Santé, et de l'initiative interdisciplinaire DEFISENS du CNRS. Elle se positionne ainsi en complémentarité des laboratoires ISIR, LTCI, L2E, IMS, INL et LIRMM en particulier.

Objectifs de résultats à 5 ans Les objectifs à 5 ans sont l'aboutissement de CYCLOPE à un démonstrateur de capsule endoscopique 3D et des projets WARM et STIVE à une plateforme, certifiée dispositif médical, de e-santé pour les images et les signaux médicaux.

5 Analyse SWOT

Points forts : – *Cohésion* : bon esprit équipe, homogénéité, membres unis

– *Réactivité* : dynamisme, adaptabilité, anticipation

– *Complémentarité scientifique* : large spectre connaissance, équilibre, qualité scientifique

– *Capacités* théorie : analyse, modélisation ; pratique : réalisation matérielle, expérimentation, applications

Points faibles : – *Ressources* : faible nombre d'équivalents-temps-plein en recherche du fait des fortes responsabilités pédagogiques, soutien de base insuffisant pour des activités en électronique

– *Structure* : peu de HDR dans l'équipe, pas d'ingénieur pour supporter réalisations et expérimentations

Risques liés au contexte : – *Marginalité* : électronique marginale au sein de l'UPMC

– *Université Paris Saclay* : émergence d'un pôle scientifique majeur concurrent

Possibilités liées au contexte : – *UPMC* : force, qualité, structures solides : UFR919, LIP6, Polytech Paris UPMC, UFR de médecine ; Labex SMART, Idex SUPER, SATT LUTECH

– *Paris - Ile de France* : localisation, richesse du tissu académique, médical (AP-HP) et industriel, structures régionales, dynamisme

– *Besoins socio-économiques* : besoins de santé publique, besoins des entreprises

– *Tendance* évolution de l'électronique vers les systèmes et rapprochement de l'informatique

6 Mise en œuvre

Valorisation et recherche de moyens : La valorisation des brevets déposés par l'UPMC dans le domaine médical s'appuiera sur la création d'une startup, BIONICOM. Les moyens seront recherchés avec des par-

tenaires industriels ou académiques selon la maturité des projets (TRL).

Diffusion des résultats (stratégie de publication) : L'équipe continuera à privilégier les revues internationales dans sa politique de publication. Ce sont les vecteurs qui apportent la plus grande visibilité et les évaluations les plus enrichissantes des travaux. Elle participera davantage aux conférences de référence que dans la période de référence si ses ressources le permettent. Cette stratégie de publications soutiendra le soutien de HDR par les Maîtres de Conférences de l'équipe, ce qui rééquilibrera sa structure.

Stratégie locale / nationale/ Internationale : La stratégie locale s'appuiera sur les axes transverses du LIP6 et le Labex SMART, auquel participent en particulier l'ISIR, le L2E et le LTCI. La stratégie régionale s'appuie sur la construction de relations durables avec des partenaires industriels, comme CIRA (projet WARM en cours et autres projets à venir comme STIVE), NXP (projet ANR WiNoCoD en cours de démarrage) et THALES Avionics (projet CORAIL AME des Initiatives d'Excellence, démarrage prévu Q3 2013), en s'appuyant sur l'animation du GIS eSys (en particulier ETIS) et les pôles de compétitivité (SYSTEMATIC et CAP DIGITAL principalement). La stratégie nationale s'appuie sur le GDR SOC/SIP (dans lequel l'équipe a de fortes responsabilités dont la direction) et le CNRS sur le plan académique, avec des actions inter-disciplinaires comme DEFISENS ; le recrutement d'un chercheur CNRS est d'ailleurs vital pour l'équipe. Sur le plan international, des relations durables sont établies avec l'Université de Gênes et les Universités de Tokyo et de Tohoku, tandis que de nouvelles relations sont en construction en Europe (Université Libre de Bruxelles), en Chine et en Afrique (Cameroun).

Implication future dans l'enseignement : L'équipe continuera à prendre des responsabilités au sein de la Licence SDI, sur la partie électronique, et de la spécialité SESI du Master en Informatique, au sein desquels elle soutient depuis une dizaine d'années l'émergence d'une filière en « Electrical and Computer Engineering ». Elle continuera aussi à prendre des responsabilités au sein de Polytech Paris UPMC, en particulier les spécialités ELI et E2I. Enfin elle soutiendra le développement d'échanges internationaux, en particulier avec l'Italie (dans le prolongement des accords Erasmus avec l'Université de Gênes que l'équipe a promus depuis quinze ans), l'Irlande (Université de Cork), les Etats-Unis, le Japon et la Chine (NWPU à Xi'An).



Lip 6



UPMC
SORBONNE UNIVERSITÉS