

# DOSSIER D'ÉVALUATION

## Vague E : campagne d'évaluation 2018-2019

### SAMOVAR UMR 5157

Services répartis, Architectures,  
Modélisation, Validation,  
Administration des Réseaux

Directeur pour le contrat en cours : Walid Ben-Ameur

Directeur pour le contrat à venir : François Desbouvries

Septembre 2018

# Table des matières

<b>1</b>	<b>Laboratoire SAMOVAR : bilan</b>	<b>5</b>
1.1	Présentation du laboratoire . . . . .	5
1.1.1	Introduction . . . . .	5
1.1.2	Effectifs et moyens du laboratoire . . . . .	6
1.1.3	Politique scientifique . . . . .	7
1.2	Présentation de l'écosystème recherche de SAMOVAR . . . . .	10
1.2.1	Animation scientifique . . . . .	11
1.2.2	Formation par la recherche . . . . .	11
1.2.3	Valorisation et Innovation . . . . .	12
1.3	Produits et activités de recherche du laboratoire . . . . .	12
1.3.1	Bilan scientifique . . . . .	12
1.3.2	Faits marquants du laboratoire . . . . .	16
1.4	Organisation et vie du laboratoire . . . . .	17
1.4.1	Pilotage, animation, organisation . . . . .	17
1.4.2	Parité, intégrité scientifique, protection et sécurité . . . . .	19
1.5	Analyse SWOT du laboratoire . . . . .	21
<b>2</b>	<b>Stratégie scientifique globale de l'unité au prochain quinquennat</b>	<b>22</b>
2.1	Stratégie scientifique . . . . .	22
2.1.1	Projet scientifique global . . . . .	22
2.1.2	Contexte institutionnel . . . . .	23
2.2	Projets des équipes . . . . .	24
2.2.1	ACMES . . . . .	24
2.2.2	ARMEDIA . . . . .	24
2.2.3	METHODES . . . . .	25
2.2.4	R3S . . . . .	25
2.2.5	TIPIC . . . . .	26
2.3	Actions spécifiques "Science des données et Intelligence artificielle" et "Internet des Objets" . . . . .	26
2.3.1	Choix des thèmes des actions spécifiques . . . . .	26
2.3.2	Projet scientifique de l'action spécifique "Sciences des données et intelligence artificielle" . . . . .	27
2.3.3	Projet scientifique de l'action spécifique "Internet des objets" . . . . .	28
<b>3</b>	<b>Équipe ACMES</b>	<b>31</b>
3.1	Présentation de l'équipe ACMES . . . . .	31
3.1.1	Introduction . . . . .	31
3.1.2	Effectifs et moyens de l'équipe . . . . .	31
3.1.3	Profils d'activité . . . . .	33
3.2	Produits de la recherche et activités de la recherche de l'équipe ACMES . . . . .	33
3.2.1	Bilan scientifique . . . . .	33
3.2.2	Faits marquants de l'équipe ACMES . . . . .	40

3.3	Organisation et vie de l'équipe ACMES . . . . .	41
3.4	Analyse SWOT de l'équipe ACMES . . . . .	42
3.5	Projet scientifique de l'équipe ACMES . . . . .	42
3.5.1	Parallélisme et système . . . . .	43
3.5.2	Génie logiciel et intergiciel . . . . .	44
3.5.3	Données et Connaissances . . . . .	46
<b>4</b>	<b>Équipe ARMEDIA</b>	<b>49</b>
4.1	Présentation de l'équipe ARMEDIA . . . . .	49
4.1.1	Introduction . . . . .	49
4.1.2	Effectifs et moyens de l'équipe . . . . .	49
4.1.3	Profils d'activité . . . . .	50
4.2	Produits de la recherche et activités de la recherche de l'équipe ARMEDIA . . . . .	50
4.2.1	Bilan scientifique . . . . .	50
4.2.2	Faits marquants de l'équipe ARMEDIA . . . . .	56
4.3	Organisation et vie de l'équipe ARMEDIA . . . . .	56
4.4	Analyse SWOT de l'équipe ARMEDIA . . . . .	56
4.5	Projet scientifique de l'équipe ARMEDIA . . . . .	57
<b>5</b>	<b>Équipe METHODES</b>	<b>61</b>
5.1	Présentation de l'équipe METHODES . . . . .	61
5.1.1	Introduction . . . . .	61
5.1.2	Effectifs et moyens de l'équipe . . . . .	61
5.1.3	Profils d'activité . . . . .	62
5.2	Produits de la recherche et activités de la recherche de l'équipe METHODES . . . . .	63
5.2.1	Bilan scientifique . . . . .	63
5.2.2	Faits marquants de l'équipe METHODES . . . . .	68
5.3	Organisation et vie de l'équipe METHODES . . . . .	68
5.4	Analyse SWOT de l'équipe METHODES . . . . .	68
5.5	Projet scientifique de l'équipe METHODES . . . . .	69
5.5.1	Projet groupe Optimisation . . . . .	69
5.5.2	Projet groupe Méthodes formelles . . . . .	70
5.5.3	Projet groupe Performance . . . . .	71
<b>6</b>	<b>Équipe R3S</b>	<b>72</b>
6.1	Présentation de l'équipe R3S . . . . .	72
6.1.1	Introduction . . . . .	72
6.1.2	Effectifs et moyens de l'équipe . . . . .	72
6.1.3	Profils d'activité . . . . .	74
6.2	Produits de la recherche et activités de la recherche de l'équipe R3S . . . . .	75
6.2.1	Bilan scientifique . . . . .	75
6.2.2	Faits marquants de l'équipe R3S . . . . .	79
6.3	Organisation et vie de l'équipe R3S . . . . .	80
6.4	Analyse SWOT de l'équipe R3S . . . . .	80
6.5	Projet scientifique de l'équipe R3S . . . . .	81
6.5.1	Optimisation de ressources pour des services sous contraintes . . . . .	82
6.5.2	Sécurité et protection des données personnelles dans le cas d'applications critiques . . . . .	83
<b>7</b>	<b>Équipe TIPIC</b>	<b>85</b>
7.1	Présentation de l'équipe TIPIC . . . . .	85
7.1.1	Introduction . . . . .	85

7.1.2	Effectifs et moyens de l'équipe . . . . .	85
7.1.3	Profils d'activité . . . . .	86
7.2	Produits de la recherche et activités de la recherche de l'équipe TIPIC . . . . .	87
7.2.1	Bilan scientifique . . . . .	87
7.2.2	Faits marquants de l'équipe TIPIC . . . . .	91
7.3	Organisation et vie de l'équipe TIPIC . . . . .	92
7.4	Analyse SWOT de l'équipe TIPIC . . . . .	92
7.5	Projet scientifique de l'équipe TIPIC . . . . .	93

# 1 Laboratoire SAMOVAR : bilan

## 1.1 Présentation du laboratoire

### 1.1.1 Introduction

#### Historique

Créé comme Formation de Recherche en Evolution (FRE 2310) au 1<sup>er</sup> janvier 2001, le laboratoire SAMOVAR est devenu l'Unité Mixte de Recherche (UMR 5157) commune au Centre National de la Recherche Scientifique et à Télécom SudParis au 1<sup>er</sup> janvier 2003. Son statut a depuis été renouvelé. Ses effectifs ont régulièrement augmenté passant de 26 à 79 enseignants-chercheurs par intégration d'enseignants-chercheurs de Télécom SudParis et de l'ENSIIE. Ce laboratoire a un rattachement principal à l'institut des sciences de l'information et de leur interactions (INS2I) et secondaire à l'institut des sciences de l'ingénierie et des systèmes (INSIS).

#### Localisation

Le laboratoire est localisé sur deux sites : Evry et Palaiseau, distants d'une trentaine de kilomètres. Une quinzaine d'enseignants-chercheurs travaillent au centre Nano-INNOV à Palaiseau. Le reste des membres est localisé à Evry (Télécom SudParis et ENSIIE). Les membres de l'ENSIIE disposent d'un bureau à Télécom SudParis. Des bureaux sont également mis à disposition des enseignants-chercheurs de Nano-INNOV lorsqu'ils se rendent à Evry. Une navette quotidienne gratuite relie les deux sites.

#### Structuration du laboratoire

Le laboratoire se structure en cinq équipes de recherche :

- L'équipe ACMES (Algorithmes, Composants, Modèles Et Services pour l'informatique répartie) comprend 17 membres permanents dont 4 professeurs, 4 directeurs d'études et 9 maîtres de conférences.
- L'équipe ARMEDIA (Applied Research for Multimedia Enrichment, Diffusion, Interaction and Analysis) comprend 11 membres permanents dont 4 professeurs, 2 directeurs d'études, 1 directeur de recherche CNRS, et 4 maîtres de conférences.
- L'équipe METHODES (Méthodes et modèles pour les réseaux) comprend 15 membres permanents dont 10 professeurs et 5 maîtres de conférences.
- L'équipe R3S (Réseaux, Systèmes, Services, Sécurité) comprend 18 membres dont 7 professeurs, 8 maîtres de conférences, 2 directeurs d'études et 1 ingénieur de recherche.
- L'équipe TIPIC (Traitement de l'Information Pour Images et Communications) comprend 18 membres dont 10 professeurs et 8 maîtres de conférences.

#### Thématiques de recherche

La recherche à SAMOVAR porte essentiellement sur l'information : comment l'acquérir, la modéliser, la coder, la transmettre, l'acheminer, la propager, la compresser, la mesurer, l'évaluer, la stocker, la protéger, la représenter, la vérifier, l'analyser, la prédire, l'inférer, la détecter, la tracer, l'enrichir, l'exploiter, etc. Ces recherches se basent sur de solides socles disciplinaires que sont l'informatique, les réseaux, les mathématiques appliquées et la physique.

Plus précisément, l'équipe ACMES mène sa recherche dans les domaines suivants :

- Données et connaissances
- Parallélisme et système
- Systèmes répartis et intergiciels
- Cloud et processus métier.

L'équipe ARMEDIA travaille sur les axes suivants :

- Image, son et vision
- Modélisation, échange et enrichissement des contenus
- Interaction et interfaces adaptées.

L'équipe METHODES mène des recherches dans les domaines suivants :

- Optimisation
- Evaluation de performances dans les réseaux
- Méthodes formelles pour le test et la vérification.

Les activités de recherche de l'équipe R3S portent sur les thèmes suivants :

- Internet des objets
- Réseaux auto-organisés
- Architectures de service et réseaux virtuels
- Sécurité et protection des données.

Enfin, les activités de l'équipe TIPIC protent sur les thèmes suivants :

- Inférence bayésienne dans des modèles markoviens
- Traitements statistiques de signaux
- Génération, transport, contrôle, détection et traitement de signaux optiques
- Localisation indoor
- Algorithmes rapides pour l'électromagnétisme.

Ces thématiques de recherche seront détaillées dans les prochains chapitres qui traiteront du bilan et du projet de chacune des équipes.

### 1.1.2 Effectifs et moyens du laboratoire

Le nombre de permanents du laboratoire a augmenté en passant de 56 en 2013 à 79 au 1/06/2018. Cette augmentation s'est faite d'une part grâce à l'intégration d'enseignants-chercheurs de l'ENSIIE ainsi que d'une équipe de TSP (Télécom SudParis) qui faisait partie d'une autre UMR (voir la rubrique Réorganisation du laboratoire du paragraphe 1.1.3). D'autre part, de nouveaux recrutements ont été faits à TSP et à l'ENSIIE et quelques EC (enseignants-chercheurs) de TSP qui n'avaient pas dans le passé une activité de recherche soutenue ont également été intégrés lorsque leur activité a atteint le niveau requis (voir 1.1.3). L'unité comprend aujourd'hui 4 EC de l'ENSIIE, 1 DR CNRS et 74 EC de TSP. Le tableau suivant indique le nombre d'EC, le nombre d'HDR (EC habilités à diriger les recherches) et le nombre de doctorants à différentes dates.

Indicateurs / dates	1/10/2005	1/10/2009	1/9/2013	1/6/2018
Enseignants-chercheurs	33	51	56	79
Docteurs d'état ou HDR	11	21	34	50
Doctorants	73	97	112	101

Outre l'augmentation du nombre d'EC que nous venons de commenter, le nombre de HDR a aussi augmenté. Notons cependant que la proportion d'HDR n'a presque pas changé entre 2013 et 2018 (autour de 62%). Le taux d'HDR reste donc élevé. Ceci s'explique en partie par la politique manageriale de l'IMT (Institut Mines-Télécom) qui encourage les EC à passer une HDR.

Notons également une baisse du nombre de doctorants entre 2013 et 2018 malgré l'augmentation du nombre de permanents et du nombre de HDR. En effet, le volume des ressources propres a connu une certaine stagnation. Notons que ceci permet d'avoir un meilleur niveau d'encadrement (en moyenne 2 doctorants par HDR, et 1.3 doctorants par EC à SAMOVAR).

Concernant le nombre de soutenances, nous comptons en moyenne 30 thèses soutenues par an. La durée moyenne des thèses est de 41 mois. Environ 280 thèses se sont déroulées en totalité ou en partie pendant ce quinquennat (thèses qui ont commencé après janvier 2013 ou thèses soutenues après janvier 2013).

Concernant le mode de financement de ces thèses, la figure 1.1 donne la répartition des thèses selon le type de financement entre janvier 2013 et juin 2018. Nous considérons ici 6 types : CIFRE ; thèses bénéficiant d'une bourse

de l'IMT ou de TSP (y-compris celles provenant des financements CARNOT); les co-tutelles (avec l'étranger); thèses du type externe (soit le doctorant est salarié en France hors IMT, soit il est encadré par un EC de SAMOVAR sans être inscrit à TSP); les thèses du type étranger (doctorant bénéficiant d'un financement étranger); et enfin les thèses financées sur contrat (contrat de recherche).

Nous observons que 15% des thèses sont du type IMT-TSP-CARNOT. Ceci montre que nous disposons en interne de ressources permettant de soutenir des thématiques particulières. L'essentiel des thèses sont financées sur contrat (de l'ordre de 34%). Environ une thèse sur 5 est une CIFRE.

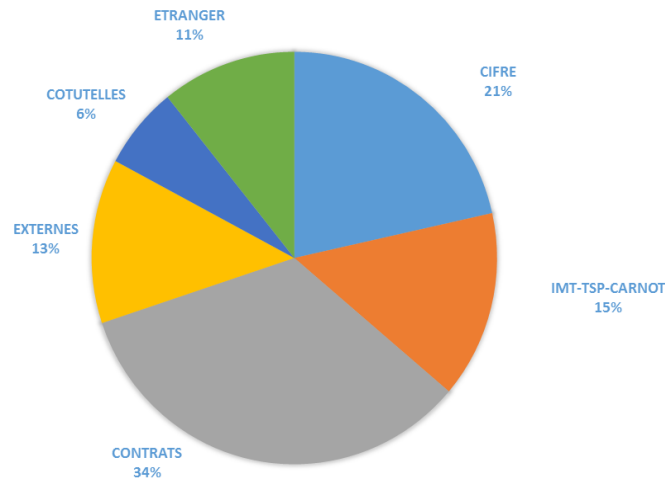


FIGURE 1.1 – Financement des thèses

En plus des EC et des doctorants, au moins deux autres catégories de personnels participent à la recherche du laboratoire : les post-doc (post-doctorants), et le personnel support-recherche (ingénieurs de recherche, ingénieurs plateformes). Pendant ce quinquennat, nous avons reçu 69 post-doctorants et 171 personnes en support-recherche. Le nombre total d'hommes-mois du type post-doc est de l'ordre de 790 alors que le nombre d'hommes-mois du type support-recherche est de l'ordre de 1360. Ainsi un post-doc reste en moyenne 11 mois à SAMOVAR alors qu'un personnel support-recherche passe en moyenne 8 mois au laboratoire.

Le secrétariat et la gestion administrative et financière de l'unité sont assurés par une ITA CNRS en plus de 6 chargées de gestion salariées de TSP. Au moins deux ingénieurs permanents réseau et plateforme contribuent aux activités du laboratoire. Il faut toutefois rajouter qu'étant hébergé par TSP et l'ENSIIE, SAMOVAR bénéficie de tout le support des deux entités impliquant ainsi beaucoup plus de personnel administratif et technique.

Les post-doctorants, les supports-recherche ainsi qu'une bonne partie de nos doctorants sont financés grâce à nos ressources propres (contrats + Carnot). Comme le montre le fichier excel présentant les données du contrat en cours (onglet Ressources Financières), les ressources de l'unité sont globalement stables aux alentours de 5 M€. Observons cependant que comme le nombre de permanents a augmenté pendant ce quinquennat, le chiffre d'affaires par EC a baissé. Ceci est principalement dû à la baisse des taux d'acceptation de certains appels à projets (ANR notamment).

Environ 50% de nos ressources propres proviennent des programmes européens. Le reste des ressources est assez diversifié. On peut noter une stabilité au niveau des contrats industriels et une croissance des investissements issus des PIA (plan d'investissement d'avenir).

### 1.1.3 Politique scientifique

Outre les actions entreprises au niveau des 5 équipes de recherche du laboratoire pour mener à bien leur propre projet scientifique, une politique est conduite au niveau du laboratoire pour soutenir les équipes, renforcer les liens

entre elles, maintenir, voire améliorer le niveau de la production scientifique.

Nous décrivons dans la suite d'une manière très concrète l'essentiel de ce qui a été entrepris au niveau du laboratoire en prenant en compte les recommandations du précédent rapport d'évaluation.

## **Actions entreprises par l'unité**

### Séminaires transversaux

Un budget spécifique a été alloué pour organiser des séminaires transversaux qui ont vocation à intéresser au moins deux équipes. François Trahay est chargé de la collecte des propositions des équipes et de la sélection des séminaires les plus pertinents.

### Événements fédérateurs

En plus des séminaires inter-équipes sus-mentionnés, une journée des doctorants est organisée tous les ans. Un ensemble de doctorants présentent leurs travaux et un jury décerne des prix en fin de journée. Cet événement très apprécié des membres du laboratoire crée une certaine émulation et contribue à faire découvrir aux uns les recherches menées par les autres. Un à deux chercheurs seniors sont également invités à cette occasion à donner une conférence (par exemple Erol Gelenbe, Serge Abiteboul, etc.).

### Soutien aux thématiques de recherche transverses

Les thématiques d'énergie et de systèmes complexes ont été identifiées dans notre précédent rapport d'évaluation comme thématiques transverses aux équipes. Il s'agit donc de thématiques qui ne sont pas propres à une seule équipe et qui ont été jugées suffisamment importantes pour être soutenues par le laboratoire.

Un animateur a été choisi pour chacune de ces deux thématiques (Marc Girod-Genet pour l'énergie et Vincent Gauthier pour les systèmes complexes). Un budget a été alloué pour l'organisation de séminaires liés à ces thématiques. Le laboratoire finance également une dizaine de bourses de stage de master tous les ans. Les stages liés à ces thématiques transverses sont prioritaires. Signalons également que deux bourses de thèse ont été accordées (financement Carnot) pour financer des travaux sur ces deux thématiques. Le bilan relatif à ces thématiques transverses est présenté au paragraphe 1.3.1.

### Maintien et amélioration de la qualité des publications

En plus des actions précédemment mentionnées qui contribuent indirectement au maintien et à l'amélioration de la qualité de la recherche réalisée au sein du laboratoire, des mesures spécifiques ont été prises pour y parvenir. D'une part, il a été décidé de financer prioritairement les missions des EC à des conférences de référence. Les publications dans des conférences moins bien classées (c'est notamment le cas des nouvelles conférences) sont financées sur les contrats.

D'autre part, le laboratoire a fait le choix de n'accueillir que des EC actifs en recherche. L'activité est évaluée et suivie via les publications sur une fenêtre temporelle de 4 ans. La "qualité" des publications est évaluée via deux bases de référence (ISI et ERA CORE). Quelques indicateurs seront présentés en 1.3.1.

### Séjours invités

La DRFD (direction de la recherche et des formations doctorales) de TSP et le laboratoire financent conjointement des séjours de chercheurs étrangers. Ces chercheurs séjournent en moyenne 1 mois à SAMOVAR. Nous finançons en moyenne 6 à 7 visites par an pour un budget total de l'ordre de 30 k€. Nous avons par exemple invité Alexandre Brandwajn de l'UC Santa Cruz, Yuriy Shamily de l'universidad de Guanajuato, Nina Evtushenko de l'université de Tomsk, Ahmed Kamal de l'Iowa State University, Marc Frappier de l'université de Sherbrooke, Jan Mendling de la Vienna University of Economics and Business, Richard Waldinger du SRI's Artificial Intelligence Center de Californie, etc.

### Soutien aux doctorants

L'unité apporte un soutien aux doctorants en finançant la participation aux écoles thématiques. En moyenne, 7 à 8 doctorants bénéficient tous les ans d'un tel soutien. Le laboratoire complète généralement le financement accordé par les écoles doctorales (STIC et Hadamard) de l'université Paris-Saclay.



### Renforcement du rôle du laboratoire dans les instances locales

Un des objectifs du quinquennat a été de renforcer le rôle du laboratoire dans les instances de TSP, plaçant ainsi la recherche au cœur des préoccupations de TSP.

Le directeur de l'unité fait désormais partie du comité de direction de TSP et participe à ce titre aux prises de décision. Le laboratoire définit également les thématiques de recherche prioritaires qui sont ensuite soumises au comité de direction de TSP qui définit les postes ouverts au concours chaque année en prenant en compte également les besoins en enseignement formulés par la direction de la formation. Notons également que pour chaque poste ouvert, le laboratoire participe au recrutement à divers titres. Le responsable de l'équipe concernée par le recrutement participe à un comité de présélection qui étudie les candidatures reçues et en sélectionne 3 à 6. Ensuite une commission de recrutement à laquelle participe le directeur de l'unité ou son représentant auditionne et classe les candidats présélectionnés.

Le laboratoire donne également un avis sur les demandes de promotion (maître de conférences vers professeur) à TSP. Plus de détails sont donnés en 1.4.1.

### Réorganisation du laboratoire

La réorganisation du laboratoire succinctement évoquée dans le projet du laboratoire en 2013, est devenue nécessaire suite à la volonté de SAMOVAR d'intégrer de nouveaux membres pour préparer son évolution dans le cadre du pôle Paris-Saclay. Il s'agissait tout d'abord d'intégrer les enseignants-chercheurs de TSP qui étaient membres de l'UMR MAP5 (rattachée à l'INSMI) depuis janvier 2010, en tant qu'équipe "Ingénierie Multimédia". Les activités de recherche de cette équipe concernent la modélisation et l'analyse des contenus numériques et sont proches des activités du groupe apprentissage statistique et reconnaissance des formes de l'équipe TIPIC (dans son ancienne configuration). L'intégration de ces EC au sein de SAMOVAR répondait également au souhait de TSP de regrouper l'essentiel de ses forces de recherche au sein d'une même UMR faisant partie de l'ensemble Paris-Saclay. La tutelle souhaitait en effet que tous les doctorants soient désormais inscrits à une même école doctorale, nécessitant ainsi l'appartenance des directeurs de thèse à l'un des laboratoires de l'ensemble Paris-Saclay. L'équipe ARMEDIA intégrant les deux groupes de chercheurs sus-mentionnés est ainsi née.

D'autre part, SAMOVAR souhaitait aussi intégrer des enseignants-chercheurs de l'ENSIIE, anciennement membres du CEDRIC (CNAM) effectuant des recherches dans l'équipe "optimisation combinatoire" et l'équipe "conception et programmation raisonnée". Ces deux domaines sont présents au sein de l'équipe METHODES de SAMOVAR. L'ENSIIE, qui est située à Evry et associée à l'Institut Mines-Télécom, souhaitait faire de SAMOVAR le laboratoire de rattachement de ses futurs recrutés dans les domaines précédemment cités. Là encore, l'intégration répondait à une volonté commune de TSP et de l'ENSIIE de se rapprocher pour créer un pôle de recherche et d'enseignement fort sur Evry.

L'intérêt pour SAMOVAR d'une telle intégration était donc double : devenir le seul laboratoire de rattachement des enseignants-chercheurs de TSP et affirmer son attractivité d'une part au niveau local (Evry) et d'autre part au niveau du pôle Paris-Saclay. La restructuration des équipes associée à cette intégration a permis une meilleure visibilité de certaines thématiques du laboratoire (il s'agissait notamment faire apparaître plus clairement l'axe "analyse de contenu numérique", l'axe "couche physique des systèmes de communication" ainsi que les axes "méthodes formelles" et "optimisation") et faciliter les synergies au sein d'une même équipe du fait d'une plus grande proximité des thèmes traités. Le résultat de la réorganisation a été présenté dans la section 1.1.1, alors que les thématiques de recherche de chaque équipe seront détaillées dans les chapitres suivants. Signalons enfin que cette réorganisation a été approuvée par les instances du CNRS, de TSP et de l'ENSIIE.

### Renforcement des liens avec les partenaires

En plus des initiatives individuelles des enseignants-chercheurs ainsi que des actions menées par les équipes pour renforcer les liens avec nos partenaires d'Evry et du plateau de Saclay (voir 1.2), le laboratoire a également joué un rôle de facilitateur. Quelques actions concrètes ont été mises en place. D'une part la réorganisation du laboratoire a permis d'intégrer des enseignants-chercheurs de l'ENSIIE comme indiqué plus haut. D'autre part, le laboratoire a pu dégager un budget de 50 k€ en 2018 pour soutenir des projets de recherche avec le laboratoire LTCI de Télécom ParisTech. En effet, Samovar et le LTCI seront partiellement co-localisés à Saclay dans un même bâtiment de l'IMT. Un appel commun aux 2 laboratoires (budget total annuel de 100 k€) a été lancé cette année, permettant ainsi de financer 7 projets communs. Cette initiative devrait être renouvelée dans le futur. Il convient également de signaler que plusieurs membres cadres du laboratoire se sont investis dans différentes structures de

coordination de la recherche faisant partie de l'écosystème recherche de l'unité. Quelques éléments sont présentés d'une manière non exhaustive dans le paragraphe 1.2. Enfin, comme indiqué précédemment, le laboratoire a œuvré à l'amplification de ses collaborations internationales en participant à l'invitation de chercheurs étrangers.

### Profils d'activité

Notre laboratoire participe aux quatre principales missions de production de la connaissance (communications dans des revues et conférences, activités concourant au rayonnement et à l'attractivité scientifiques...), de valorisation et de transfert (expertise, transfert, contrats, dépôt de brevet, interactions avec l'environnement social, économique et culturel, ...), d'appui à la communauté (animation scientifique aux niveaux national et international, GDR, organisation de conférences, participation à des comités éditoriaux, commissions, ...) et enfin de formation par la recherche (organisation et formation au niveau master, qu'il s'agisse de masters universitaires ou de Voies d'Approfondissement dans le cadre du cursus ingénieur à TSP, encadrement de doctorants et de stagiaires...).

Nous estimons que notre profil d'activité se répartit globalement comme suit.

<i>activité scientifique</i>	<i>valorisation et Transfert</i>	<i>appui à la communauté</i>	<i>formation par la recherche</i>
45 %	25 %	15 %	15 %

SAMOVAR a un profil d'activité caractérisé par une recherche académique forte associée à un appui à la communauté significatif et une activité de valorisation et de transfert importante. Du fait de notre domaine de recherche autour de l'information, pour lequel il existe un secteur industriel très fort en France, nous avons des collaborations importantes avec les acteurs industriels et des activités de valorisation riches (14 brevets déposés sur la période). Comme indiqué en 1.1.2, l'essentiel des thèses de doctorat sont financés grâce à nos contrats, nos conventions CIFRE et nos divers contacts avec les industriels. Le niveau d'investissement du laboratoire dans la formation par la recherche est également important et se matérialise par la gestion de plusieurs parcours de masters de l'Université Paris-Saclay, la participation aux enseignements de plusieurs masters, en plus des voies d'approfondissement de TSP, et l'encadrement de nos nombreux stagiaires et doctorants.

## 1.2 Présentation de l'écosystème recherche de SAMOVAR

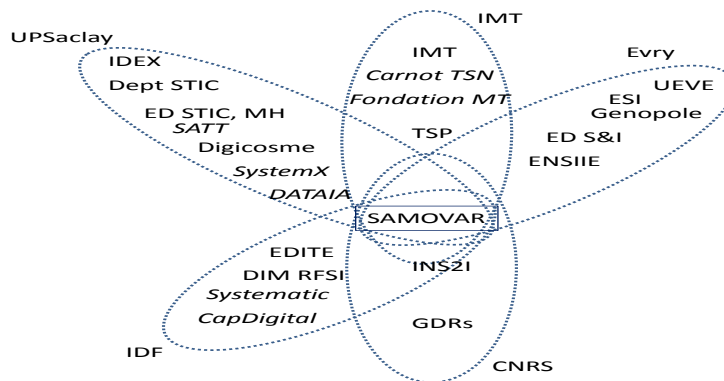


FIGURE 1.2 – Ecosystème de SAMOVAR

L'écosystème de SAMOVAR (voir Figure 1.2) se structure autour de ses deux tutelles (CNRS et Institut Mines-Telecom - Telecom SudParis) et des logiques de site (Evry, plateau de Saclay et Île de France). On peut également l'analyser selon les services supportés, à savoir l'animation scientifique, la formation par la recherche ainsi que l'innovation et la valorisation de la recherche.

Ces différentes structures s'articulent les unes par rapport aux autres et sont en général complémentaires, même si certaines fonctions peuvent être assurées à différents niveaux (la valorisation par exemple).

### 1.2.1 Animation scientifique

L'animation et la politique scientifique du laboratoire s'appuient en partie sur notre écosystème. Nous sommes actifs dans un certain nombre de GDRs du CNRS (ISIS, GPL, MADICS, RO, RSD, pré GDR Sécurité, ...). En particulier, François Desbouvries est un des directeurs scientifiques adjoints du GDR ISIS, et Catherine Dubois est une des responsables du pôle Langages et Vérification du GDR GPL. Quelques projets du type PICS ont été soutenus par le CNRS, dont le projet "Réseaux petit monde en Mouvement" en collaboration avec des partenaires de Singapour.

Dans l'Université Paris-Saclay, nous sommes membres du département STIC et partie prenante d'une IRS (Initiative de Recherche Stratégique) B5GI (Beyond 5G Initiative), qui regroupe les forces transverses du laboratoire dans ce domaine. Djamal Zeglache est membre du conseil du département STIC. Nous bénéficions également de l'animation de la fondation de mathématiques Jacques Hadamard. Walid Ben-Ameur participe au conseil scientifique de l'un de ses programmes : PGM0 (programme Gaspard Monge pour l'optimisation). Le laboratoire participe également à l'institut de convergence DATAIA regroupant les acteurs de Saclay autour des sciences des données et de l'intelligence artificielle. SAMOVAR participe notamment au projet PEPER, financé par DATAIA, portant sur la prédiction par apprentissage de la prosommation d'énergie renouvelable.

SAMOVAR participe également au labex Digicosme qui fédère la recherche du plateau de Saclay autour des thématiques Communications, Science du logiciel et Science des données. Le labex anime la recherche via des groupes de travail et nous sommes actifs notamment dans des groupes autour de la thématique réseaux. Maryline Laurent co-anime un axe sécurité dans le labex, Bruno Defude co-anime la commission formation et fait partie du comité exécutif, et Michel Marot a animé le groupe de travail "Performances, QoS et gestion de ressources des réseaux sans fil dynamiques".

Au sein de l'IMT nous avons participé activement à l'animation scientifique inter-écoles structurée par les réseaux thématiques (RT) du domaine télécom. En particulier nous avons co-animé les RT réseaux, multimédia, sécurité. Au delà des réunions scientifiques régulières, c'est au travers de ces RT que des financements de thèses (1 à 2 par an) ont été possibles. Plus récemment l'IMT, du fait de l'intégration des écoles des Mines, a préféré regrouper les forces de recherche et valorisation sur deux programmes transverses en santé et Industrie du Futur grâce auxquels nous bénéficions de financement légers par des fonds d'amorçage. Nos activités de recherche en santé, cybersécurité, internet des objets entre autres ont été largement valorisées lors des colloques scientifiques bi-annuels de l'IMT. Dans sa stratégie 2018-2022, l'IMT poursuit sa structuration dans ce sens au travers de la définition de 9 thématiques phares à fort potentiel recherche et valorisation, associées à des transformations sociétales (numérique, énergie, santé, mobilité, etc.). La stratégie de SAMOVAR au regard de cette nouvelle structuration de la Recherche et Innovation de l'IMT sera évoquée dans la partie projet.

Le laboratoire fait partie du réseau Francilien en Sciences Informatiques (RFSI), labelisé en décembre 2016 par la région Île-de-France en tant que Domaine d'Intérêt Majeur (DIM) émergent. Le DIM RFSI fédère l'ensemble des acteurs franciliens de la recherche en sciences et technologies de l'information et de la communication en s'appuyant sur un réseau de 27 laboratoires partenaires. Le projet NumAMMa (NUMA Memory Manager) porté par SAMOVAR a été financé par le DIM RFSI.

Au niveau d'Evry, l'animation se fait principalement via l'association Evry Sciences et Innovation (ESI) qui regroupe tous les acteurs de l'enseignement supérieur et de la recherche de la communauté d'agglomération. ESI organise de l'ordre de 3 séminaires de recherche thématiques chaque année dont l'objectif est de mettre en relation les acteurs locaux afin de susciter des collaborations (SAMOVAR a par exemple animé un séminaire sur les big data et un autre sur les objets connectés). Un colloque scientifique de plus grande ampleur est organisé par ESI chaque année avec pour objectif de montrer à l'extérieur les résultats des travaux effectués sur le site. Les plateformes du laboratoire sont en partie hébergées dans le bâtiment ETOILE situé sur le campus de TSP et qui héberge également les actions en matière d'innovation et de valorisation (incubateur, fablab). Evry est pour nous également porteur d'une dynamique autour de la santé avec la présence d'acteurs importants (Génopole, AFM Genethon, Centre Hospitalier Sud Francilien), ce qui nous a permis la création d'un living lab (EVIDENT) et d'une plateforme IoT pour la santé (Hadaptic).

### 1.2.2 Formation par la recherche

Du côté de la formation par la recherche (master et doctorat), pendant ce quinquennat nous avons basculé progressivement vers l'université Paris-Saclay. Initialement nos doctorants étaient inscrits soit dans l'ED (école

doctorale) Sciences et Ingénierie (co-accréditation UEVE et TSP) pour l'équipe ACMES, soit dans l'ED EDITE (co-accréditation UPMC, Telecom ParisTech, TSP, CNAM) pour les autres équipes. Nous avons déployé tous nos programmes de master en co-accréditation vers les masters et parcours de Saclay et nous avons également basculé tous nos doctorants inscrits dans l'ED S&I d'Evry vers Saclay (école doctorale STIC) et nous avons inscrit les nouveaux doctorants très majoritairement dans l'école doctorale STIC et quelques uns dans l'ED Mathématiques Hadamard. Les doctorants qui ont commencé leur thèse à EDITE ont continué dans cette école (ils auront tous terminé fin 2018).

Nos doctorants sont tous financés grâce à nos interactions avec notre écosystème (les types de financement obtenus sont exposés en 1.1.2). Nous avons également obtenu quelques bourses de master (en prévision d'une thèse) via l'IDEX Paris Saclay et le labex Digicosme.

### 1.2.3 Valorisation et Innovation

La valorisation et plus largement les relations industrielles sont importantes dans notre unité en grande partie parce que notre tutelle IMT/TSP dépend du ministère de l'industrie et que les enseignants-chercheurs de TSP ont une mission de valorisation en plus de leurs missions d'enseignement et de recherche. Ces trois missions ne sont d'ailleurs pas antagonistes mais peuvent tout à fait se renforcer les unes les autres (une relation industrielle démarrée via un stage d'élève ingénieur peut se poursuivre par une thèse CIFRE ou un contrat de recherche bilatéral par exemple). SAMOVAR collabore régulièrement avec plusieurs grandes entreprises dont Orange, Nokia (Alcatel-Lucent), EDF et Thalès. De plus, l'institut Carnot TSN (Télécom et Société Numérique), géré par l'IMT, nous assure un abondement sur les contrats industriels que nous pouvons réinvestir soit en ressourcement (contrats doctoraux), soit en valorisation (ingénieurs plateforme ou développement). Nous sommes également présents dans d'autres instruments de valorisation comme l'IRT SystemX (diverses thèses dans les domaines data, mobilité et transport), la SATT Paris Saclay (à noter le projet CBORBIG, avec le CNRS) et bien sûr les deux pôles de compétitivité de notre domaine (CapDigital et Systematic). Nous sommes d'ailleurs très actifs dans Systematic en étant présents dans plusieurs groupes de travail et avec des responsabilités importantes dans les groupes Telecom (Djamal Zeghlache), Smart Cities (Titus Zaharia) et Confiance Numérique, Sécurité et défense (Hervé Debar).

## 1.3 Produits et activités de recherche du laboratoire

### 1.3.1 Bilan scientifique

Nous commençons par commenter le bilan global de l'unité avant de présenter brièvement le bilan associé aux thématiques transverses.

#### Bilan global de l'unité

Le bilan de l'unité étant l'addition des bilans des équipes, les détails scientifiques seront exposés aux prochains chapitres. Nous nous contenterons donc ici de commenter quelques indicateurs liés à la production scientifique de SAMOVAR.

Pendant ce quinquennat, 167 thèses ont été soutenues et 14 brevets ont été déposés. L'unité a publié 504 articles de revue, 1017 articles de conférence, 34 chapitres de livre et 3 livres.

Même s'il est très difficile d'évaluer la qualité d'une publication, nous avons décidé de nous référer à la base australienne ERA CORE qui classe les publications en plusieurs catégories (A\*, A, B, C), les publications non-classées étant considérées comme de moins bonne qualité. Le classement des conférences date de 2018 alors que celui des journaux date de 2010. Ainsi tous les nouveaux journaux créés après 2010 ne sont pas classés. Il convient aussi d'indiquer que la base ERA CORE référence mal les publications dans certaines disciplines comme la physique (notamment les publications du groupe optique de l'équipe TIPIC). Le tableau suivant présente les nombres de publications de type A\* et A.

	Revue A*	Revue A	Conf A*	Conf A
Nombre	52	112	32	130

Nous observons que non seulement le volume global de nos publications est important, mais le laboratoire a réussi à publier dans les meilleurs supports (selon ERA CORE) d'une manière régulière et soutenue. Ainsi, en considérant

que l'effectif moyen sur le quinquennat était de 69.5, un EC a été auteur ou co-auteur d'approximativement 8.6 articles de revue, et 17.6 articles de conférences<sup>1</sup>. Si l'on divise par la durée de la période considérée (5.5 années), on en déduit qu'un EC de SAMOVAR est chaque année co-auteur de 1.56 articles de revue, et 3.2 articles de conférence.

Pour avoir un autre éclairage de la production scientifique de l'unité, nous avons utilisé la base SCOPUS et observé certaines métriques propres à cette base. Nous résumons ici les résultats constatés. Le nombre total de publications référencées dans SCOPUS sur la période 2013-2018 (juin) s'élève à 1651. Le nombre de citations des articles publiés pendant cette période est de 7036. Ainsi, le nombre moyen de citations par article est de l'ordre de 4.3. Si on tient compte de la moyenne habituelle de citations selon les domaines (comme proposé par SCOPUS), l'impact moyen de nos publications est de 1.38 (tout rapport supérieur à 1 indique une performance meilleure que la moyenne du domaine). Enfin, 15.5% de nos publications sont parues dans les 10% des supports les plus cités (conférences et journaux).<sup>2</sup>

## Thématiques transverses

Le laboratoire avait identifié dans son projet l'énergie et les systèmes complexes comme thématiques transverses aux équipes. Comme indiqué précédemment, le laboratoire a pu dégager des moyens pour soutenir ces initiatives. Nous présentons brièvement le bilan de nos efforts sur ces deux thématiques.

### A. Énergie

La mise en avant de cette thématique transverse est justifiée par les forces des équipes de SAMOVAR sur les technologies de base applicables au "domaine Énergie". Les enjeux ainsi que le positionnement des équipes de SAMOVAR et leurs principaux travaux et résultats dans cette thématique sont résumés ci-après.

Depuis ces dernières années, le domaine de l'énergie est en pleine mutation. Cette mutation est le fait de la libéralisation du marché de l'énergie en Europe avec l'apparition de producteurs d'énergie indépendants et décentralisés, et la volonté affichée de l'Europe de réduire d'un facteur 4 les émissions de CO2 d'ici 2050. Cette décentralisation des systèmes et du marché de l'énergie s'opère en parallèle de la transition énergétique qui impose à la fois une minimisation de l'impact de la production d'énergie sur l'environnement (baisse du nucléaire et introduction de flexibilité, i.e. augmentation des énergies renouvelables) et une meilleure gestion du ratio entre production et consommation. L'introduction de la flexibilité et l'apparition de la mobilité électrique rendent ces tâches plus complexes. En effet :

- certaines productions d'énergies renouvelables, comme par exemple le solaire ou l'éolien, sont intermittentes (e.g. production et rendement d'une éolienne moindre si trop ou peu, voire pas de vent),
- les consommateurs commencent à disposer dans leurs habitats de solutions de production renouvelables et de micro-stockage sur batteries. Ils deviennent ainsi des "prosumers",
- la mobilité électrique et l'usage de véhicules électriques sont en augmentation : lieu et niveau de consommation d'énergie deviennent moins facilement évaluables et prévisibles.

Sachant que les opérateurs ne disposent toujours pas de moyens viables de stockage d'énergie électrique à grande échelle, ces nouvelles contraintes et caractéristiques des réseaux électriques complexifient la gestion d'efficacité énergétique et en augmentent le coût.

Les travaux de SAMOVAR portant sur l'énergie peuvent être classés en deux ensembles : les travaux autour des Télécoms vertes, et ceux autour des réseaux énergétiques du futur. Nous présentons brièvement les contributions du laboratoire sur chacun des deux thèmes.

Les axes impliquant des EC de SAMOVAR (équipes ACMES, METHODES, R3S) et concernant les Télécoms Vertes sont les suivants :

- réduction de la consommation d'énergie de terminaux mobiles et d'équipements embarqués (e.g. Smartphone, BANs, WSNs hybrides),

---

1. Nous avons naturellement pris en compte le fait qu'une publication pouvait avoir plusieurs auteurs qui sont membres permanents du laboratoire

2. Signalons que dans tout le reste du rapport, les références citées dans le texte font nécessairement partie de la sélection des publications indiquée dans les annexes 4 des équipes (les 20%). Ainsi, la citation 2.J.ac indiquerait le deuxième article de type journal de l'équipe ACMES ("ac" pour ACMES, "ar" pour ARMEDIA, "mt" pour METHODES, "rs" pour R3S et "tp" pour TIPIC), alors que le troisième article de conférence de l'équipe METHODES est cité sous la forme 3.C.mt ("C" pour conférence, "J" pour journal et "L" pour chapitre de livre).

- amélioration de l'efficacité énergétique des réseaux d'accès/d'agrégation et des Datacenters /clouds (e.g. stockage et allocation de ressources, routeurs optiques, réseaux vidéo IP, Small Cells, HetNets),
- introduction d'énergies renouvelables (e.g. panneaux solaires) et de batteries dans des réseaux pour stocker/utiliser/revendre de l'énergie via une Smart Grid.

Les apports de SAMOVAR concernant les axes susmentionnés ont été les suivants :

- évaluation/modélisation/classification de comportements d'utilisation et de consommation : d'équipements, de terminaux, de ressources, de systèmes, de technologies de transmission (réseaux cellulaires, réseaux sans fil hybrides et courte portée), de protocoles, de réseaux,
- évaluation de performances pour des réseaux économes en énergie [6.C.rs, 23.C.rs, 43.C.rs, 66.C.rs, 31.C.ac],
- spécification et validation de nouveaux algorithmes/mécanismes réseau énergie-efficace (transmission/routage, allocation de ressources, scheduling, handover, stockage de données, traitements distribués, CDN, SDN, datacenter/cloud/terminaux optiques/stations de base) [1.C.mt, 16.C.mt, 22.J.rs, 16.C.tp, 33.C.ac].

Les axes impliquant des EC de SAMOVAR (équipes ACMES + R3S + METHODES + TIPIC) et concernant les réseaux énergétiques du futur sont principalement les suivants :

- gestion d'efficacité énergétique dans un contexte d'une production d'énergie décentralisée et de flexibilité,
- modélisation de Smart Grids, de profils énergétiques et d'utilisateurs,
- facturation dynamique,
- cyber-sécurisation des protocoles SCADA,
- description d'un système énergétique et de ses données (incluant mesure, contrôle, management, business).

Les apports de SAMOVAR concernant les axes susmentionnés ont été les suivants :

- modélisation de Prosumers agrégés [19.C.mt, 32.C.mt],
- modélisation d'un profil énergétique d'un territoire (incluant modèle météo/producteurs/ consommateurs),
- design de mécanismes d'intégration et de gestion de micro-stockage pour un réseau plus stable,
- design de mécanismes d'agrégation locale de production d'énergie plus efficace (i.e. moins volatile) [9.C.mt],
- mécanismes de contrôle d'un réseau de Prosumers avec meilleure gestion de phase, minimisation de l'énergie de contrôle et maximisation de la résilience.
- modèle de connaissance sémantique ouvert et ontologies modulaires pour la description d'un système énergétique et de ses données, avec gestion d'interopérabilité sémantique [23.J.rs].

Pour réaliser ces travaux, les EC de SAMOVAR se sont notamment appuyés sur les méthodes et outils suivants : programmation dynamique, MDP (Markov Decision Processes), théorie des graphes, optimisation, modélisation de données et sémantique, fouille et analyse de données (Data Analytics), systèmes d'apprentissages, systèmes multi-agents collaboratifs,

Les projets de recherche en cours sur contrats des équipes de SAMOVAR sur l'axe de recherche "Télécom vertes" (Green IT) sont les suivants :

- européens : H2020 SILENCE (ultra Sound Interfaces and Low Energy iNtegrated SEnsors), Celtic+ CONVINCe (Consumption OptimizatiON in VIdéo NEtworks), Celtic+ SOOGREEN (Service-oriented optimization of Green mobile networks),
- nationaux : ANR N-GREEN (New-Generation of Routers for Energy Efficient Networks), Digicosme PERFECO (Evaluation de performances pour réseaux économes en énergie), PGMO BOBIBEE (Bornes et Bid Prices dans le But d'Economiser l'Énergie).

Les projets de recherche en cours sur contrats des équipes de SAMOVAR sur l'axe de recherche "réseaux énergétiques du futur" sont les suivants : ITEA3 SEAS (Smart Energy Aware Systems, avec pour objectif de permettre à l'ensemble des acteurs du marché de l'électricité d'interagir en temps réel avec les systèmes de production et de consommation afin d'optimiser la consommation globale d'énergie), ITEA3 FUSE-IT (Future Unified System for Energy and Information Technology, pour répondre au besoin d'un système de gestion des bâtiments durable, fiable, convivial, efficace, sûr et sécurisé dans le contexte de sites critiques intelligents),

Le chiffre d'affaires lié à ces activités contractuelles de SAMOVAR est donné dans le tableau 1.1.

2013	2014	2015	2016	2017	2018
31 k€	198 k€	503 k€	560 k€	562 k€	345 k€

TABLE 1.1 – Chiffre d'affaires des activités contractuelles de SAMOVAR dans la thématique Énergie

Le tableau 1.2 donne le nombre de publications ainsi que les thèses soutenues ou en cours en lien avec la

thématique énergie.

Période	Livres	Journaux	Conférences	Rapports/Spécifications	Thèses en cours/ soutenues
2013-2018	1	24	48	5	11/13

TABLE 1.2 – Publications et thèses de SAMOVAR dans la thématique Énergie

Notons enfin qu'une journée dédiée à l'énergie a été organisée le 26 septembre 2017 à SAMOVAR. En plus de membres de SAMOVAR, plusieurs chercheurs extérieurs ont été invités à faire des exposés. Le programme est consultable sur <http://SAMOVAR.telecom-sudparis.eu/spip.php?article1053>. Le laboratoire a également participé au 8<sup>ème</sup> colloque scientifique de l'IMT du 28 avril 2017 sur le thème "Énergie en révolution numérique".

## B. Systèmes complexes

On appelle systèmes complexes des systèmes grands et dynamiques. L'identification des systèmes complexes comme thématique transverse avait pour but de stimuler les collaborations interdisciplinaires et multi-équipes autour des problématiques suivantes : le passage à l'échelle des grands systèmes, la dynamique des grands systèmes, et les systèmes en interaction. Les réseaux de transport sont également très dynamiques. Le cloud peut être vu comme un système complexe constitué de sous-systèmes hétérogènes. Nous présentons brièvement quelques résultats obtenus selon plusieurs axes.

### B.1. Mobile Crowdsensing (MCS)

L'activité de recherche sur le mobile Crowdsensing est le fruit d'une collaboration entre les équipes ACMES pour les aspects algorithmiques et R3S pour les aspects relatifs aux télécoms. Durant le dernier quinquennat, un ensemble de méthodes a été développé afin d'optimiser la collecte d'informations provenant des capteurs intégrés dans les téléphones mobiles dans un contexte urbain. Les méthodes proposées ont pour but d'optimiser la collecte sous diverses contraintes : d'énergie, de qualité de données, de recrutement des participants [29.J.rs, 38.C.ac, 33.C.ac, 6.C.ac, 22.J.ac]. Cette activité a été financée au travers d'un projet H2020 Monica et 2 thèses ont été soutenues sur le sujet. La production scientifique liée à cette activité a été particulièrement prolifique.

### B.2. Clustering et Méthodes d'agrégation

Sous cet axe de recherche nous avons regroupé deux groupes de travaux avec différents domaines d'applications, mais qui ont pour point commun l'utilisation de méthodes liées soit aux algorithmes de détection de communautés pour les graphes, soit aux méthodes d'agrégation appliquées aux théories des files d'attente. Cet axe est le fruit d'une collaboration entre les équipes METHODES et R3S.

- La première approche concerne les méthodes de détection de communautés au travers des problèmes de k-séparateurs. Nous caractérisons plusieurs classes de graphes pour lesquels le problème se résout polynomialement, nous proposons des algorithmes d'approximation, et nous développons des approches polyedrales pour le problème [46.C.mt, 7.J.mt].
- La deuxième approche concerne la modélisation des relations contractuelles dans le cadre d'un marché de l'électricité avec les acteurs (producteurs et consommateurs appelés "prosumers"). Il a été développé une méthode de détection de communautés dans les graphes (méthode de percolation de cliques) pour former des coalitions optimales de "prosumers" dans le contexte du smart grid [3.J.mt, 19.C.mt, 32.C.mt]. Cet axe a été financé par un projet ANR Marmote, ainsi qu'une bourse de l'IMT pour un total de quatre thèses sur cette sous-thématique.
- Le troisième travail porte sur l'analyse des performances et la consommation énergétique dans le Cloud. Nous avons utilisé pour cela des modèles de files d'attente avec seuils, de type hystérésis. Ce type de modèles de file d'attente permet de représenter la dynamique des ressources en fonction de la charge [6.J.mt, 15.J.mt].

### B.3. Graphes de connaissances et analyse sémantique

Cet axe de recherche est le fruit d'une collaboration entre les équipes ARMEDIA et ACMES qui a permis le développement de mécanismes de reconnaissance d'activités humaines basés sur les technologies sémantiques (ontologies, données liées, RDF, etc.). Les algorithmes de reconnaissance d'activités ont été testés avec les données provenant des capteurs installés dans le Living-Lab EVIDENT. Le living-Lab Evident est un appartement test permettant d'explorer et d'évaluer les usages des technologies de la santé, du numérique et de la dépendance. Les méthodes développées dans ce cadre prennent en compte des imperfections des données de capteurs, la gestion d'événements complexes (CEP) et de plusieurs dimensions de l'incertitude (fraîcheur, consistance, précision et contradiction) [3.C.ac, 14.C.ac, 15.C.ac, 12.J.ac]. Cette activité a été financée au travers d'un projet ANR CAIR (2015-2018) et d'un projet FUI CoCAPS (2016-2018). Quatre thèses ont été financées sur cette thématique.

#### B.4. Dynamique d'opinion et dissémination de l'information dans les réseaux sociaux

Dans cet axe thématique des méthodes ont été développées pour étudier la dynamique d'opinion et la dissémination d'information dans les réseaux sociaux. Les équipes METHODES, ACMES et R3S ont participé à cet axe. Pour l'étude de la convergence d'opinions nous avons utilisé un modèle de Degroot étendu pour étudier la compétition d'opinions par exemple dans un contexte électoral. Cette étude a fait appel à l'optimisation et la théorie des jeux et a donné lieu à plusieurs publications dont [3.C.mt]. Une autre étude de la propagation d'information a été menée en prenant en compte les thématiques des informations diffusées modélisées au travers d'un processus de Hawkes [42.C.mt]. Cet axe a été financé par le projet ANR Idefix, et trois thèses ont été soutenues sur le sujet dont une financée par une bourse de l'IMT. Un post-doc a été financé sur le sujet.

#### B.5. Analyse de la dynamique de la Mobilité

Cet axe de recherche est le fruit d'une collaboration entre les équipes R3S et ARMEDIA. Les équipes ont développé des mécanismes qui permettent l'analyse de la mobilité humaine au travers des données de signalisation de téléphonie mobile. Nous avons proposé CT-Mapper, un algorithme non supervisé qui permet la cartographie des traces de téléphonie mobile sur un réseau de transport multimodal. L'une des principales forces de CT-Mapper est sa capacité à cartographier les trajectoires multimodales cellulaires avec des données clairsemées sur un réseau de transport multicouche où les couches ont des propriétés physiques différentes. Un tel réseau est modélisé par un grand graphe multicouche dans lequel les nœuds correspondent à des stations de métro/train ou à des intersections de route et les arêtes correspondent à des connexions entre eux. Le problème de mapping est modélisé par un HMM (modèle de Markov caché) non supervisé où les observations correspondent à des trajectoires mobiles d'utilisateurs clairsemés et les états cachés aux nœuds de graphes multicouches [21.J.rs, 28.J.rs, 31.C.rs]. Dans le cadre de ce travail 3 thèses ont été soutenues, 1 post-doc a été financé. Cet axe a été financé par un projet FUI Fluidtrack.

En résumé, le nombre de thèses liées à la thématique transverse "systèmes complexes" était de l'ordre de 12, dont 2 financées par des bourses IMT et 10 sur fonds propres. Deux post-doctorants ont également été recrutés. La thématique a été soutenue par 1 projet H2020, 3 projets ANR, et 2 projet FUI.

### 1.3.2 Faits marquants du laboratoire

Nous énumérons ici quelques faits qui ont marqué la vie du laboratoire pendant le dernier quinquennat.

- Le laboratoire a été réorganisé et ses effectifs ont augmenté d'une manière sensible, lui permettant ainsi de renforcer son rôle d'acteur important de la recherche dans les sciences de l'information. La réorganisation et l'intégration de nouveaux membres se sont faites en maintenant la cohérence thématique du laboratoire autour des sciences de l'information.
- La participation de l'IMT et de Télécom SudParis à la création de l'université Paris-Saclay a été un élément marquant pour le laboratoire. En effet, le laboratoire est devenu de fait l'un des laboratoires de l'UPSAclay et a participé à certaines de ses instances de gouvernance. Des financements ont pu être obtenus grâce à notre appartenance à l'UPSAclay et des collaborations avec d'autres laboratoires du plateau ont été initiées ou renforcées. La création plus récente de "NewUni" qui comprend la plupart des écoles d'ingénieurs qui étaient dans UPSaclay commence à impacter la vie du laboratoire (séminaires communs, création en cours de la faculté "data science and information technologies" à laquelle SAMOVAR appartiendra, établissement de bonnes bases pour une future stratégie scientifique commune). L'installation d'une quinzaine de membres



du laboratoire à NanoInnov (Saclay) pendant ce quinquennat a évidemment permis de tisser encore plus de liens avec les acteurs de la recherche à Saclay.

- Le laboratoire a bénéficié d’une chaire de recherche CNRS autour de l’optimisation distribuée marquant ainsi la confiance du CNRS et contribuant à dynamiser encore plus la recherche au sein de l’unité. Des chaires industrielles ont également été créées avec d’autres partenaires dont une portant sur les valeurs et politiques des informations personnelles et une autre sur la cybersécurité.
- Des membres de l’unité ont organisé des conférences internationales importantes dont ICSOC’2014, RAID’2016, Tests&Proofs’2018 et DIMVA’2018.
- Le laboratoire participe à plusieurs projets importants (et parfois les pilote) dont les projets européens EUNITY, CSA, SILENCE, MONICA, reTHINK, CONVINCe, Wise-IoT, SUPERCLOUD, etc.
- Plusieurs prix de meilleurs articles ont été obtenus dans des conférences dont UbiComp’2015, UbiComp’2016, UIC’2015, MobiQuitous’2017, RCIS’2016, SCC’2018, CAiSE’2018, APCC’2017, IWSEC’2016, CPSCoM’2013 et PICOM’2017. Une thèse à SAMOVAR a également reçu le prix DGA de la meilleure thèse en 2016.
- Plusieurs plateformes de recherche ont pu être créées ou consolidées pendant ce dernier quinquennat permettant ainsi à l’unité d’expérimenter et valider certains de ses résultats.
- Plusieurs membres du laboratoire bénéficient désormais d’une très bonne reconnaissance de la communauté scientifique. Nous pouvons par exemple citer la participation aux comités éditoriaux de plusieurs revues dont IEEE Transactions on Signal Processing, Signal Processing, Annals of applied probability, IEEE Pervasive Computing, Computers&Electrical Engineering, Computers&Security, Signal Processing : Image Communication, ACM Transactions on Intelligent Systems and Technology, etc. Nous pouvons également mentionner l’invitation de plusieurs membres à dispenser des conférences plénières dans des congrès dont GPC’2018, INFORMS Telecommunication conference’2016, ACM Turing 50th Celebration Conference’2017, UIC’2014, etc.

## 1.4 Organisation et vie du laboratoire

### 1.4.1 Pilotage, animation, organisation

#### Processus de prise de décision

##### Conseil de laboratoire

Le laboratoire est doté d’un conseil comprenant 15 membres, dont 6 enseignants-chercheurs élus, un élu représentant les ingénieurs, techniciens et personnels administratifs, un élu représentant les doctorants, 6 membres nommés et le directeur de l’unité. Le conseil a été renouvelé le 12 mars 2015. Le conseil se réunit en moyenne tous les 2 mois. Un compte-rendu est diffusé à la suite de chaque réunion du conseil. Les conseils sont généralement ouverts à tous les membres du laboratoire. Des réunions supplémentaires sont parfois organisées pour traiter de sujets particuliers : promotions, certains appels à projets, préparation de l’évaluation du laboratoire, etc.

##### Jurys pour différents appels

Une à deux bourses de thèse sont accordées tous les ans par l’IMT au laboratoire. Un jury composé des responsables d’équipes, du directeur de l’unité, de la directrice de la recherche et du directeur adjoint de la recherche de TSP classe les candidatures. Le classement est proposé à l’IMT qui accorde une à deux bourses selon la qualité des dossiers proposés par toutes les écoles de l’IMT.

Une à deux bourses de thèse sont également accordées tous les ans grâce aux budgets Carnot. Le laboratoire participe au jury qui accorde ces bourses.

Comme indiqué en 1.1.3, une dizaine de bourses de stage de master sont accordées tous les ans. Le jury comprend la directrice de la recherche de TSP, les responsables d’équipe, les chefs de département de TSP et le directeur de l’unité.

Rappelons également que le laboratoire a pu dégager un budget de 50 k€ en 2018 pour soutenir des projets de recherche avec le laboratoire LTCI de Télécom ParisTech. Un appel commun aux 2 laboratoires (budget total de 100 k€) a en effet été lancé cette année, permettant ainsi de financer 7 projets communs. Le jury était composé des membres du conseil de laboratoire.

### Avis sur les promotions

Le laboratoire donne un avis sur les demandes de promotion (maître de conférences vers professeur) à TSP. Cet avis est établi par le conseil de laboratoire en se basant sur deux rapports demandés en interne à un collègue de la même équipe et un collègue d'une autre équipe. L'avis est ensuite exprimé par le directeur du laboratoire ou son représentant pendant la commission de qualification pour promotion et changement d'appellation.

### Répartition du budget

La répartition du budget est discutée en conseil de laboratoire qui décide du montant alloué à chaque action transverse. Le reste du budget est réparti entre les équipes proportionnellement au nombre d'enseignants-chercheurs de chaque équipe.

### Règlement intérieur

L'unité dispose naturellement d'un règlement intérieur qui a été discuté lors de plusieurs réunions du conseil de laboratoire. Ce règlement décrit plusieurs procédures dont notamment les modalités de vote du conseil.

## **Animation transverse et interaction entre les équipes**

Comme indiqué précédemment, un budget spécifique a été alloué pour organiser des séminaires transversaux qui ont vocation à intéresser plus d'une équipe. Un collègue collecte les propositions des équipes et sélectionne les séminaires les plus pertinents.

En plus des séminaires inter-équipes susmentionnés, une journée doctorants est organisée tous les ans. Un ensemble de doctorants présente leurs travaux et un jury décerne des prix en fin de journée (voir 1.1.3).

Rappelons également que les thématiques d'énergie et de systèmes complexes ont été identifiées comme thématiques transverses aux équipes. Un animateur a été nommé pour chacune de ces deux thématiques. Un budget a été alloué pour l'organisation de séminaires liés à ces thématiques. Le laboratoire accorde une dizaine de bourses de stage de master tous les ans. Les stages liés aux thématiques transverses sont prioritaires. Comme indiqué précédemment, deux bourses de thèse ont été accordées (financement Carnot) pour financer des travaux sur les 2 thématiques.

## **Plateformes mutualisées**

La stratégie d'innovation de SAMOVAR est en partie construite autour des plateformes de recherche, dont l'activité permet de favoriser l'innovation industrielle. Les expérimentations réalisées dans les plateformes permettent à l'unité et aux entreprises partenaires d'aborder plus concrètement leurs sujets avant d'aller plus loin dans l'innovation.

SAMOVAR regroupe 8 plateformes. Quatre d'entre elles sont hébergées dans le bâtiment ETOILE :

- BioMica : plateforme pour l'imagerie médicale et la bio imagerie
- Living-Lab Evident complété par la plateforme Hadaptic : un appartement test pour explorer et évaluer les usages des technologies de la santé, du numérique et de la dépendance, complété par une plateforme autour de l'internet des objets appliqué au domaine de la dépendance.
- THD : plateforme de recherche sur les Réseaux Très Haut Débit
- NCF : (Network Cloud Federation) plateforme expérimentale cloud et réseaux.

Quatre autres plateformes technologiques de SAMOVAR sont également présentes sur le campus d'Evry ou à NanoInnov, à Palaiseau :

- Cybersécurité : plateforme pour la sécurité des infrastructures connectées
- M4IoT : plateforme des composants logiciels pour l'Internet des Objets
- Plenimage : plateforme de microscopie haute résolution, grand champ
- MyMultimediaWorld : plateforme cloud du traitement multimédia

Le laboratoire participe également au projet SESAME Saclay-IA qui vise à rendre accessibles un ensemble de ressources matérielles (espaces d'échange et de stockage, nœuds de calcul) et de logiciels bien adaptés aux besoins spécifiques des algorithmes d'intelligence artificielle (parallélisation) et munis d'accélérateurs graphiques

(GPU). Un budget annuel de 20 k€ a été accordé à cette initiative par TSP. Signalons enfin que l'IMT co-gère la plateforme TERALAB autour des données massives. Un des membres du laboratoire est responsable scientifique de la plateforme qui est ouverte à toute la communauté.

## Communication

Le laboratoire dispose d'un site Web (<http://SAMOVAR.telecom-sudparis.eu/>) dont le contenu est régulièrement mis à jour. On y annonce les séminaires, les soutenances de thèse, certains appels à projets, etc. Plusieurs listes de diffusion propres au laboratoire (doctorants, permanents, ensemble du laboratoire, liste par équipe, etc.) sont utilisées pour diffuser les informations.

Les EC du laboratoire contribuent également à la diffusion des résultats de recherche obtenus dans le laboratoire à travers des articles de vulgarisation dans différents journaux. Citons par exemple l'article "De nouveaux outils contre Alzheimer" du 30/05/2016 dans le journal du CNRS ou encore l'article intitulé "La certification anonyme pour mieux protéger les données personnelle" paru dans les actualités scientifiques de l'INS2I du mois de septembre 2016 (<http://www.cnrs.fr/ins2i/spip.php?article2235>). Des EC du laboratoire sont régulièrement interviewés ou participent à des émissions de radio ou de télévision pour communiquer sur leur domaine de recherche (voir par exemple le reportage de France 3 sur la 5G diffusé au journal télévisé du 19/03/2018, ou encore l'émission "la méthode scientifique" de radio France culture, diffusée le 2 mai 2018).

## Politique de recrutement

Le comité de direction de TSP, où SAMOVAR est représenté, définit la politique de recrutement. Il a ainsi été décidé de recruter principalement des enseignants-chercheurs au niveau maître de conférence et rarement au niveau professeur. La charge d'enseignement des nouveaux recrutés est généralement maintenue à un faible niveau au cours des premières années pour laisser le temps aux jeunes maîtres de conférences de démarrer une activité de recherche d'un bon niveau.

Concernant le profil des EC recrutés, chaque équipe du laboratoire a défini deux thématiques de recherche prioritaires pour les futurs recrutements à TSP. Ces thématiques sont ensuite soumises au comité de direction de TSP qui définit les postes ouverts au concours chaque année en prenant en compte également les besoins en enseignement formulés par la direction de la formation.

Notons également que pour chaque poste ouvert, le laboratoire participe au recrutement à divers titres. Le responsable de l'équipe concernée par le recrutement participe à un comité de présélection qui étudie les candidatures reçues et en sélectionne 3 à 6. Ensuite une commission de recrutement, à laquelle participe le directeur de l'unité ainsi que quelques membres de l'équipe concernée par le recrutement, classe les candidats présélectionnés.

## Participation à des labex, equipex, Carnot, etc.

Comme exposé en 1.2, le laboratoire participe à plusieurs structures qui sont représentées dans la figure 1.2. Cette participation nous a permis d'interagir avec la plupart des acteurs de la recherche du domaine STIC de l'Île-de-France.

### 1.4.2 Parité, intégrité scientifique, protection et sécurité

#### Parité

Signalons tout d'abord que le taux de féminisation dans le corps des enseignants-chercheurs s'élève à 29% alors qu'il atteint 38% pour les doctorants. Observons également que parmi les 13 enseignants-chercheurs membres du conseil de laboratoire, 5 sont des femmes. Le représentant des Ingénieurs, Techniciens et personnels Administratifs est également une femme. Bien que ces taux soient importants, nous n'avons pas encore atteint le niveau de parité souhaité. Le laboratoire essaye dans la mesure du possible de favoriser le recrutement de femmes. Des actions précises ont été initiées sur le campus. Nous pouvons en citer quelques unes.

- Une formation/sensibilisation contre le sexisme des membres du comité de direction de TSP a été organisée le 14 mai dernier.
- Une référente égalité a été nommée. Elle tient une permanence une fois par mois pour discuter du sujet.

- TSP est membre actif de "Elles Bougent" (deux étudiantes de TSP ont obtenu un prix à l'égalité pour un projet lors du Challenge Innovatech organisé en décembre par "Elles Bougent").
- Trois cours de Sciences humaines (1 sur les stéréotypes de genre, 1 sur le management de la diversité et 1 sur les rapports entre genre et technologie) sont offerts en option à l'ensemble des étudiants de TSP.
- Une campagne d'affichage "pas de sexisme dans nos jurys d'admission" a été menée.
- Un MOOC "Mixité dans le monde du numérique" est en cours de création.
- Un groupe de travail "information, prévention et accompagnement sur le Harcèlement sexuel et les violences sexistes" est en cours de constitution.
- TSP participe au groupe de travail "égalité" de la CGE (conférence des grandes écoles) où l'école est représentée par sa référente égalité.
- Notons enfin qu'un plan d'action égalité, précis et ambitieux, a été défini pour le campus d'Evry. Les actions à mener seront donc amplifiées dans les années à venir.

### **Intégrité scientifique**

Le laboratoire s'est appuyé sur les écoles doctorales qui proposent des formations liées à l'intégrité scientifique à l'intention des doctorants.

Le conseil de laboratoire est naturellement l'instance qui traite de cette question. C'est le conseil qui pourrait, par exemple, instruire des dossiers du type plagiat. Notons cependant que le problème ne s'est pas posé pendant ce quinquennat.

Nous sommes également très vigilants par rapport aux conflits d'intérêt qui peuvent se poser lorsque des décisions doivent être prises : jury de bourse de thèse, jury de bourses de master, rapports pour évaluer des dossiers de collègues demandant une promotion, etc.

### **Protection et sécurité**

Une des thématiques de recherche du laboratoire porte sur la protection des données. Le laboratoire est donc très sensible à cette problématique. La tutelle IMT a pris la mesure de l'importance de la protection des données personnelles en mettant en place plusieurs actions. Depuis la rentrée en vigueur du RGPD (règlement général sur la protection des données), le correspondant Informatique et Libertés à l'IMT est devenu DPO (Data protection officer ou délégué à la protection des données). Il a de nombreuses missions qu'il a commencé à exercer :

- Il participe à la sensibilisation de tous les services de l'IMT (dont TSP) à la protection des données. Ceci se fait à travers des rencontres avec les différents services, des réunions, des communications générales à tout le personnel, etc.
- Il participe également à l'information des usagers de l'IMT de leurs droits (par exemple en rajoutant des mentions légales sur les sites web ou les formulaires que certains sont amenés à renseigner).
- Il identifie les données personnelles à l'IMT et recense tous les traitements qu'elles subissent, détermine la finalité de chaque traitement et tient un registre contenant toutes ces informations.
- Il vérifie que les consentements ont bien été obtenus (et stockés) pour le recueil et le traitement de toutes les données privées.
- Il vérifie que les données ne sont pas conservées au delà de la durée nécessaire au regard des finalités pour lesquelles elles sont traitées.
- Il vérifie qu'une étude d'impact sur la vie privée a été faite avant la mise en place de tout nouveau traitement.
- Il vérifie que toutes les mesures nécessaires ont bien été prises pour assurer la protection des données récoltées (mesures de sécurité de tout type).
- Il alerte les responsables des traitements et rend compte à la direction de l'IMT de toute entorse au RGPD.
- Il assure le contact avec la CNIL (commission nationale de l'informatique et des libertés).

Pour mener à bien sa mission, le DPO bénéficie de formations assurées par la CNIL ainsi que de moyens mis à disposition par l'IMT. Notons qu'un référent Informatique et Libertés a également été nommé dans chacune des écoles de l'IMT (dont TSP). Ces référents travaillent en étroite collaboration avec le DPO pour l'aider dans sa mission. Le laboratoire est évidemment concerné par ces actions puisque des données personnelles sont parfois traitées dans quelques unes des recherches que nous menons.

## 1.5 Analyse SWOT du laboratoire

L'analyse de nos activités, de nos résultats et de notre positionnement conduit à l'auto-évaluation suivante :

Points forts	<ul style="list-style-type: none"> <li>— Un haut niveau de publications avec une moyenne annuelle de 1.56 articles de revues et 3.2 articles de conférence par enseignant-chercheur.</li> <li>— Une qualité des publications en constante amélioration avec 162 articles publiés pendant ce quinquennat dans des conférences de rang A ou A* selon ERA CORE-2018 et 164 articles dans des journaux de rang A ou A* selon ERA CORE-2010.</li> <li>— Un fort engagement dans les activités de valorisation et de transfert se traduisant par de nombreux projets de tous types et des collaborations avec les industriels de référence.</li> <li>— Des recherches très variées mais toutes centrées autour de l'information assurant ainsi une cohérence globale et une couverture large des sciences de l'information.</li> <li>— Un niveau du rayonnement scientifique du laboratoire en hausse comme le montrent la participation aux comités scientifiques des conférences et comités éditoriaux de revues, les différentes distinctions et l'animation ou participation à différents groupes de travail et sociétés savantes.</li> <li>— Une forte interaction avec l'écosystème du laboratoire.</li> <li>— Une hausse notable de la participation du laboratoire dans les prises de décision au niveau de TSP, notamment en ce qui concerne les recrutements et la recherche.</li> </ul>
Points à améliorer	<ul style="list-style-type: none"> <li>— Même si le laboratoire est associé à la plupart des décisions qui concernent l'affectation des budgets de recherche à TSP, les moyens propres dont dispose la direction du laboratoire pour appliquer sa stratégie scientifique restent modestes.</li> <li>— Malgré d'excellents recrutements réalisés pendant ce quinquennat (7 recrutements en 2018 à TSP dont un CR CNRS recruté comme professeur), le laboratoire a encore des difficultés à recruter des chercheurs CNRS.</li> </ul>
Opportunités	<ul style="list-style-type: none"> <li>— Le futur déménagement partiel du laboratoire vers le plateau de Saclay (2019-2020) aidera le laboratoire à se rapprocher d'acteurs académiques avec qui nous avons déjà des contacts et des collaborations (LIX, CMAP, CREST, LTCl, LRI, L2S, LIMSI, INRIA, CEA, etc.). Ceci facilitera encore plus les collaborations et permettra au laboratoire de travailler dans un cadre scientifique encore plus favorable.</li> <li>— La création de l'université "New Uni" et de sa faculté "data science and information technologies" à laquelle Samovar devrait appartenir, constitue une formidable opportunité pour envisager des projets scientifiques de plus grande envergure avec une visibilité nationale et internationale plus importante.</li> <li>— TSP vient de recruter 7 nouveaux permanents qui devraient rejoindre le laboratoire et contribuer au renouvellement de ses membres.</li> </ul>
Risques	<ul style="list-style-type: none"> <li>— Le déménagement partiel à Saclay pourrait créer des difficultés liées à la bi-localisation de certaines équipes de recherche. Le travail en commun entre les membres d'une même équipe pourrait en pâtir. Le laboratoire devra faire le nécessaire pour minimiser ce risque.</li> <li>— Le financement de nos recherches se fait beaucoup grâce aux contrats industriels et projets collaboratifs de tous types. Même si le laboratoire a réussi à renouveler ces financements, ils restent incertains et très compétitifs.</li> </ul>

## 2 Stratégie scientifique globale de l'unité au prochain quinquennat

### 2.1 Stratégie scientifique

#### 2.1.1 Projet scientifique global

En 2013 le projet du laboratoire était constitué des projets des équipes, auxquels s'ajoutaient deux projets transverses, l'un dévolu aux systèmes complexes, le second à l'énergie. Pour le prochain quinquennat le projet du laboratoire sera également la réunion des projets des équipes, renforcés par deux actions spécifiques (AS) : Sciences des données et Intelligence artificielle (SDIA) ; Internet des objets (IdO). Cette évolution s'explique notamment par les facteurs développés ci-dessous.

##### *Évolution des équipes*

Les projets des équipes reflètent l'évolution naturelle des thématiques, mais aussi du périmètre de SAMOVAR. De 2013 à 2018, le laboratoire s'est enrichi de 23 nouveaux membres (enseignants-chercheurs (EC) de TSP qui ont rejoint SAMOVAR ; recrutements TSP ; EC de l'ENSIIEE ; EC de TSP précédemment dans MAP5) et s'est restructuré en 5 équipes : ACMES, ARMEDIA, METHODES, R3S, TIPIC (ARMEDIA n'existant pas en 2013). L'activité "parallélisme et systèmes" d'ACMES n'existait pas en 2013 ; ARMEDIA est constitué essentiellement d'EC qui ont quitté TIPIC et d'anciens membres de l'UMR MAP5 ; les EC de l'ENSIIEE ont considérablement renforcé les axes "optimisation" et "méthodes formelles" de METHODES ; les EC en technologies avancées pour les communications ont quitté R3S pour rejoindre TIPIC.

##### *Devenir des thématiques transverses "Énergie" et "Systèmes complexes"*

Les activités en cours dans ces deux thèmes vont naturellement se poursuivre, mais elles se retrouvent, au moins partiellement, dans les deux AS SDIA et IdO.

En ce qui concerne le thème "Énergie", les études dans l'axe "Green IT" vont viser, d'une part à améliorer l'efficacité énergétique des réseaux, d'autre part à réduire la consommation énergétique de terminaux mobiles et équipements embarqués. Concernant le second axe "Réseaux énergétiques du futur", les thématiques de recherche devraient évoluer et s'inscrire désormais dans le cadre d'un centre de recherche inter facultés de NewUni, dédié au changement climatique et à la transition énergétique. Ce centre, en cours de concrétisation, et au développement duquel SAMOVAR participe activement, s'articulera autour de trois axes : "captage et stockage de l'énergie" ; "bâtiment intelligent" ; "prospective énergétique dans un monde connecté". Le laboratoire SAMOVAR sera essentiellement concerné par les deux derniers axes, et participera à des projets multidisciplinaires impliquant d'autres laboratoires de NewUni (le Laboratoire de Météorologie Dynamique notamment) ou du Plateau de Saclay (GeePS de CentraleSupélec).

Les études sur les systèmes complexes ont vocation à se poursuivre dans le cadre des deux AS SDIA et IdO. En effet, d'une part l'IdO constitue de par sa grande taille et sa dynamique un système complexe qui devrait être étudié. D'autre part, les données massives liées aux systèmes complexes devraient être traitées et analysées par différents outils d'intelligence artificielle et de sciences des données. Les travaux du laboratoire sur le "crowdsensing", ou encore l'analyse des réseaux sociaux ou la dynamique d'opinions seront poursuivis et renforcés.

##### *Rôle des AS dans le projet, bénéfices attendus et politique de soutien*

Les deux AS sont pensées comme des actions ciblées qui contribueront à nourrir le projet global du laboratoire, i) par un accent mis sur des thématiques pertinentes, ii) par l'exploitation et la mise en valeur de nos compétences propres et iii) par la consolidation de certains points encore fragiles grâce à des recrutements ciblés de qualité (nous gagnerions à nous renforcer dans le domaine hautement concurrentiel de la SDIA, où nous sommes encore de jeunes acteurs). Elles contribueront par ailleurs à fédérer plus globalement les efforts du laboratoire : les thématiques SDIA et IdO étant transverses à l'ensemble des équipes, des séminaires ciblés sur ces deux AS devraient, en suscitant l'intérêt d'un public élargi, contribuer à renforcer la vie scientifique globale de SAMOVAR et constituer ainsi une plus-value par rapport aux séminaires et travaux de recherche propres aux équipes.

Le laboratoire aura donc une politique de soutien à ces AS, qui se manifestera par l'accueil de professeurs invités, le soutien à des manifestations collectives d'intérêt (workshops, écoles d'été), et l'attribution de bourses de stage et de thèse.

## 2.1.2 Contexte institutionnel

Ces choix s'inscrivent pleinement dans les stratégies de recherche des deux tutelles de notre laboratoire (l'IMT et le CNRS), ainsi que dans la stratégie en cours d'élaboration du nouvel institut "NewUni" auquel nous participons. Enfin l'écosystème régional (Institut Convergence DATAIA, entreprises du plateau de Saclay, futurs Instituts interdisciplinaires d'intelligence artificielle (3IA)) est également bénéfique au projet.

### *Tutelles : IMT et CNRS*

Une des vocations de l'IMT est d'être un acteur majeur de la transition numérique. Cette ambition fait donc partie de sa stratégie de recherche, laquelle est déclinée ensuite au niveau de chaque école. TSP s'inscrit pleinement dans ce contexte et souhaite donc amplifier et valoriser ses actions pour la transition numérique, ce qui se traduit par un investissement accru dans un certain nombre de thématiques phares, dont la SDIA et l'IdO.

Le traitement et la fouille de grandes masses de données, la sécurité et la protection des informations, d'une part ; les réseaux, l'internet du futur et le calcul distribué, d'autre part, sont deux des six axes stratégiques de recherche de l'INS2I.

### *NewUni*

Au cours de l'automne 2017, le gouvernement a décidé de regrouper l'École Polytechnique, l'ENSAE, l'ENSTA, Telecom ParisTech et Telecom SudParis en un nouvel ensemble, appelé provisoirement "NewUni", ce qui implique notre sortie de l'université Paris Saclay. Cet ensemble NewUni est en cours de structuration ; en termes de recherche il devrait s'articuler en 5 facultés (*Natural Sciences and Mathematics, Engineering sciences, Data Sciences and Information Technologies, Humanities and Social Sciences, et Management Sciences*).

La faculté *Data Sciences and Information Technologies* regroupe le CMAP, le CREST, le LIX, le LTCI, l'U2IS et SAMOVAR (et, à titre secondaire, l'I3 et des unités d'HEC) ; toutes ces unités utilisent les mathématiques appliquées de façon intensive, et/ou sont actives dans le domaine des technologies de l'information. Les thématiques de recherche de NewUni ont été regroupées selon 5 grandes transitions sociétales (transition énergétique et changement climatique, transition sécuritaire, transition numérique, transition technologique et transition sanitaire). Notre projet de laboratoire s'inscrit donc essentiellement dans le cadre de la transition numérique, dont les axes majeurs sont l'intelligence artificielle ; la simulation numérique et la science des données ; les réseaux et l'internet des objets. Notons que du fait de l'importance actuelle du domaine de l'intelligence artificielle (encore soulignée récemment par le rapport Villani), NewUni pourrait voir émerger un centre interdisciplinaire en intelligence artificielle ; notons également que les acteurs du pôle Paris Saclay (NewUni et l'université Paris Saclay) unissent leurs forces pour répondre à l'appel 3IA.

À court terme, l'organisation de séminaires communs permettant aux enseignants-chercheurs de mieux se connaître puis de bénéficier d'expertises complémentaires devrait être un des premiers bénéfices concrets de ces structures nouvelles. Ces séminaires pourront être organisés à Évry comme à Saclay, éventuellement avec une logistique de visioconférences afin de pallier les difficultés dues à la bilocalisation.

### *Impact sur les Masters*

Notre appartenance à NewUni aura également des conséquences en termes d'enseignement. L'offre de formation en masters de Telecom SudParis, portée par les membres de SAMOVAR, s'inscrit aujourd'hui en totalité au sein

de l'Université Paris-Saclay. Ces masters évolueront, pour le prochain contrat quinquennal, vers des formations et diplômes du futur établissement NewUni. Relevant principalement des trois mentions "Informatique", "Mathématiques et Applications" et "Énergie Électrique, Électronique et Automatique", ils accueilleront des élèves ingénieurs (écoles membres de NewUni), des étudiants internationaux ainsi que des étudiants issus d'autres établissements français. Ces formations conduiront les diplômés soit vers une poursuite en doctorat soit vers une insertion professionnelle dans une entreprise, en général, de haute technologie. La pertinence, la qualité et l'actualité de ces programmes reposeront de façon étroite sur l'activité scientifique des laboratoires de NewUni.

## 2.2 Projets des équipes

Cette section contient un bref résumé des projets des équipes de SAMOVAR (se référer aux projets détaillés des équipes pour plus d'information).

### 2.2.1 ACMES

Le projet de l'équipe ACMES s'articule autour de trois axes : *Performance, Génie logiciel et intergiciel et Données et connaissances*, pour répondre à trois défis majeurs, à savoir la complexité des architectures matérielles et des piles logicielles, le caractère dynamique à la fois des infrastructures et des données, massives et hétérogènes, et le passage à l'échelle :

- *Performance*. Il est important d'améliorer les performances des infrastructures logicielles pour éviter de gaspiller les ressources matérielles et l'énergie. Nous nous focaliserons en particulier sur l'étude des abstractions que doit fournir un système pour rendre l'utilisation de ces architectures simples et efficaces, l'amélioration des performances et de la consommation énergétique des piles logicielles patrimoniales et enfin l'impact de la désagrégation et de la géo-distribution des données sur les architectures orientées stockage.
- *Génie logiciel et intergiciel*. La complexité, l'hétérogénéité et la dynamique inhérentes aux infrastructures matérielles et logicielles rendent les codes difficiles à écrire et à maintenir. Il devient donc nécessaire de proposer de nouveaux intergiciels, abstractions et modèles pour simplifier la conception et le développement des applications. Nous nous proposons de focaliser nos efforts selon deux points de vue : l'IdO, et les procédés métiers (notamment l'orchestration et l'élasticité des ressources Cloud). Il s'agira donc d'aborder d'une part les problèmes liés au service de découverte pour l'IdO, à la distribution d'informations de contexte en provenance de l'IdO à différentes échelles et à la protection de la vie privée et d'autre part, les problèmes liés à la modélisation des ressources Cloud et leur orchestration ainsi que la vérification formelle de propriétés comme l'élasticité.
- *Données et connaissances*. Il est important de proposer et d'étudier de nouveaux algorithmes permettant de collecter et traiter efficacement les grands volumes de données hétérogènes. En particulier, trois axes seront explorés : la relaxation des requêtes sur des flux de données sémantiques, le traitement déclaratif de grands flux de données sémantiques incertains mais aussi l'analyse en temps réel de réseaux sociaux en flux.

### 2.2.2 ARMEDIA

Le projet de l'équipe ARMEDIA s'articule autour de deux défis scientifiques sur la thématique SDIA avec plusieurs axes applicatifs en Santé et autonomie, Ville intelligente, Industrie du Futur et IdO.

Le premier défi, *Approches intégratives pour des données hétérogènes multicapteurs*, concerne le traitement conjoint d'une information variée et complémentaire (son, imagerie 2D, 3D, 3D+t multimodale, vidéo, signaux, données cliniques, démographiques, génomiques, etc.) en intégrant des modèles de raisonnement fondés sur des ontologies, la fusion de données, les approches statistiques, géométriques et les graphes, tout en s'appuyant sur les nouveaux paradigmes de l'intelligence artificielle. L'objectif est d'aboutir à une analyse discriminante de ces données (détection/reconnaissance, classification) pour une aide à la décision, avec un pouvoir prédictif relatif au suivi d'un traitement (médecine personnalisée) ou au suivi d'une activité dans un flux (multimédia). Les actions associées sur le volet applicatif portent sur l'imagerie biomédicale, l'analyse des signaux physiologiques et l'assistance à la personne dépendante, l'identification biométrique et la sécurité, la reconnaissance de gestes et le suivi d'activités dans les scènes, et l'analyse de scène pour contenus TV diffusés.

Le deuxième défi visé, *Analyse et traitement de données multimédias massives connectées*, s'inscrit dans un



contexte applicatif lié aux nouvelles technologies de l'industrie du futur et de la ville intelligente, régi par l'expansion continue des masses de données générées entre autres par les nouveaux dispositifs connectés. Lorsque ces dispositifs comportent des capacités multimédias, avec une partie de traitement embarqué des données, on parle de "internet of media things". L'exploitation de ce type de données massives et hétérogènes renvoie vers des problématiques d'analyse temps réel en environnement distribué/embarqué ciblant plusieurs applications multimédias dont les suivantes font l'objet de nos études : modélisation des codes logiciels qu'un utilisateur malveillant insère dans un autre contenu, traçabilité des contenus par des techniques blockchain, compression de contenus et traitement distribué, adaptation des contenus pour des échanges sur terminaux légers nomades, et analyse, décision et optimisation de flux d'information par apprentissage.

### 2.2.3 METHODES

Les travaux se poursuivront dans les trois axes : *Optimisation*, *Méthodes formelles* et *Évaluation de performance*.

Dans l'axe *Optimisation*, nous travaillerons sur des problèmes d'optimisation combinatoire et de programmation mathématique, comme par exemple le problème de flot concurrent maximal dans le cas multi-sources ou encore la généralisation d'un résultat de l'équipe portant sur le dominant des r-coupes disjointes. Nous étudierons aussi des problèmes de partitionnement et de couverture dans les graphes, avec des approches linéaires et semi-définies. Enfin, nous poursuivrons l'étude de l'approche multi-polaire pour l'optimisation robuste proposée récemment par l'équipe et nous évaluerons son apport dans le cadre d'applications liées aux réseaux.

Dans l'axe *Méthodes formelles*, nous continuerons nos travaux sur la vérification et la preuve formelle : nous nous concentrerons sur les applications des outils de vérification, y compris en temps réel, les problèmes liés au développement formel d'outils de détection d'intrusion, en utilisant la méthode formelle B/EventB, l'utilisation des ontologies pour la conception de systèmes complexes et critiques, et le développement et les applications de FoCaLiZe ainsi que l'interopérabilité des preuves et des prouveurs. Pour le thème test et monitoring, nous nous intéresserons au test des systèmes communicants, le monitoring distribué fonctionnel, l'interopérabilité des protocoles, le test actif en utilisant TDL (Test Description Language), et la mesure logicielle en y intégrant des techniques d'apprentissage non-supervisé. En ce qui concerne le génie logiciel formel, nous envisageons de travailler sur l'interaction des services, les méthodes formelles et l'apprentissage par problèmes, et les lignes de produits logiciels.

Dans l'axe *Performance*, nos travaux concerneront l'évaluation de performance, l'allocation de ressources et l'étude du compromis entre performance et consommation d'énergie dans les réseaux mobiles, notamment 5G (y compris le transport des services dits critiques et IdO), le Cloud et les réseaux optiques (notamment d'accès). La recherche de politiques optimales pour le fonctionnement des systèmes tout en économisant l'énergie se fera par des modèles mathématiques tels que les chaînes de Markov, les MRP (Markov Reward Process), et les MDP (Markov Decision Process), avec des notions d'apprentissage pour apprendre les paramètres de l'environnement dans lequel évolue notre système.

### 2.2.4 R3S

Le projet de R3S porte sur les deux défis suivants :

- *L'optimisation de ressources pour des services sous contraintes* : la problématique porte sur le déploiement de services virtualisés ou non, et l'ajustement continu, dynamique et rétroactif de ce déploiement pour satisfaire au mieux un ensemble de contraintes de sécurité ou de consommation de ressources. Trois cas d'usage permettent de soutenir notre démarche expérimentale : les services virtualisés, les fonctions de sécurité et les réseaux électriques. Les techniques envisagées relèvent (1) de l'algorithmique associée à des fonctions de supervision pour permettre une prise de décision éclairée quant à la migration de réseaux virtuels par exemples, et (2) de l'analyse statistique (réseaux de neurones, apprentissage par renforcement) pour prédire et anticiper les demandes de ressources, et aussi inférer d'autres informations d'intérêt comme l'identification de l'origine des dégradations de services, ou de nouveaux indicateurs d'aide à la décision.
- *La sécurité et protection des données personnelles dans le cas des applications critiques de e-santé et industrie 4.0* : ces applications s'appuient sur des systèmes communicants composés d'objets connectés (actionneurs, capteurs et contrôleurs), souvent en association avec des services cloud. Le choix spécifique de ces deux applications offre l'avantage de traiter de besoins communs de bonne résistance aux cyber-attaques, de résilience, de sûreté de fonctionnement, de conformité au Règlement général sur la protection des données

personnelles (RGPD), d'interopérabilité des dispositifs, mais elles se distinguent par des hypothèses de travail différentes - en termes d'administration plus ou moins verrouillée, de nombre d'acteurs en interaction, d'ergonomie - qui conduisent à des approches différentes. L'équipe R3S envisage de travailler sur la détection d'attaques et l'ajustement de la politique de réaction, l'isolation des attaquants dans le système, la continuité de service en mode dégradé, la personnalisation de services respectueuse de la vie privée, et la gestion dynamique de consentement en conformité avec le RGPD.

### 2.2.5 TIPIC

Le projet de l'équipe TIPIC s'articule en 3 axes : *Statistique et optimisation* ; *Détection, estimation, communication et surveillance en environnements complexes* ; *Optique et photonique*.

- Dans l'axe *Statistique et optimisation* un premier projet consiste à revisiter divers problèmes d'inférence statistique ou d'optimisation dans un contexte de données massives et/ou de grande dimension (qui se présente dans des applications telles que les données astronomiques, les flux vidéo, les grands réseaux de capteurs, les cubes d'images médicales), contexte dans lequel les techniques classiques de statistique computationnelle et d'optimisation ne passent pas à l'échelle. Un second projet consistera à explorer et exploiter les liens entre certaines structures neuronales utilisées en apprentissage statistique et certaines modélisations probabilistes markoviennes.
- Les réseaux de capteurs hétérogènes interconnectés produisent des données massives et fonctionnent dans des domaines électromagnétiques encombrés, ce qui rend nécessaire le développement de nouvelles techniques de communication, détection, estimation et surveillance qui soient fiables, robustes et fonctionnant à différentes échelles. Nous nous proposons dans ce contexte d'explorer des problèmes d'allocation automatique de capteurs, de détection et de décodage conjoint dans des réseaux, et de localisation ou de détection d'interférences dans des environnements contraints.
- Enfin les travaux de l'axe *Optique et photonique* s'articulent comme dans la période précédente autour de trois sous-axes. Les études consacrées au transport de l'information s'orienteront sur l'apport potentiel des algorithmes d'apprentissage automatique aux communications sur fibre optique et, réciproquement, à l'implémentation sur composants photoniques de tâches d'apprentissage automatique ; l'axe consacré aux capteurs photoniques se focalisera sur la microscopie sans marqueur par ptychographie de Fourier ; les travaux sur les lasers seront consacrés à la dynamique de lasers à semi-conducteurs en présence de perturbations externes.

## 2.3 Actions spécifiques "Science des données et Intelligence artificielle" et "Internet des Objets"

### 2.3.1 Choix des thèmes des actions spécifiques

Les pertinences scientifique, économique et sociétale de l'axe IdO, dont les champs applicatifs sont très variés (e-santé, ville intelligente, industrie 4.0) ne font pas de doute. Le rapport Villani a par ailleurs rappelé récemment les enjeux économiques de la science des données et de l'intelligence artificielle.

Au-delà des pertinences intrinsèques de ces thématiques, et de l'environnement institutionnel dont nous bénéficions, nous devrions pouvoir apporter des contributions scientifiques significatives dans ces domaines, du fait de nos compétences propres et d'une stratégie de recrutement.

- Dans la thématique IdO, nos forces sont importantes du fait de notre implication historique en télécommunications, et transverses, puisque nos compétences et savoir faire couvrent certains domaines du traitement du signal, de l'optique, des réseaux et de l'informatique. Nos travaux couvrent donc un large spectre incluant entre autres les réseaux de capteurs avec application à la 5G, les technologies photoniques pour le transport, la gestion des questions d'interopérabilité, de sécurité et de confiance, l'enrichissement sémantique des services et données, les problèmes d'allocation et de déploiement optimal des ressources dans le edge/cloud, les systèmes informatiques répartis, middleware et capteurs embarqués, et la virtualisation et le cloud. Nous sommes donc à même de proposer un projet qui couvre de nombreux aspects du problème (couche physique, systèmes de communication, couche réseau, aspects liés à la sécurité, intergiciel, applications).

- Nous avons dans le domaine SDIA des compétences transverses qui couvrent divers champs disciplinaires : optimisation, statistiques, apprentissage automatique, informatique distribuée et calcul à haute performance, bases de données distribuées, IA contextuelle et symbolique ; les applications en sont diverses et touchent plusieurs domaines : santé, mobilité, réseaux sociaux, multimédia, robotique. Même si ce domaine est extrêmement compétitif, nous pensons que nos savoir faire actuels pourraient nous permettre de développer cette thématique, d'une part par quelques contributions méthodologiques originales, d'autre part par quelques productions remarquables éventuellement liées à des applications spécifiques.
- Le laboratoire participe aux commissions de recrutement de TSP et cherche à améliorer encore la qualité et la pertinence de ces recrutements. En 2018 TSP vient d'effectuer plusieurs recrutements, dont un CR CNRS HDR, qui a vocation à rejoindre le laboratoire SAMOVAR sur un poste de Professeur en charge de l'animation transverse en SDIA, et un MC en IdO. Notre ambition est de renforcer et de structurer nos forces par des recrutements de très bon niveau, en bénéficiant de plus de la future politique de recrutement de NewUni (appels d'offres unifiés, vivier de qualité bénéficiant du prestige de NewUni).

### 2.3.2 Projet scientifique de l'action spécifique "Sciences des données et intelligence artificielle"

Les points méthodologiques qui seront explorés dans le cadre de cette AS seront, entre autres, les liens entre structures neuronales et modèles probabilistes, et l'hybridation entre IA symbolique et apprentissage statistique. Nous explorerons également les liens entre apprentissage et codage de source, développerons des algorithmes d'apprentissage semi-supervisé ainsi que des algorithmes d'apprentissage statistique pour les réseaux et la sécurité, et prolongerons les travaux actuels sur le traitement de données massives dans des grands graphes.

#### Contributions aux réseaux neuronaux

Malgré des performances qui améliorent l'état de l'art dans de nombreuses applications, les modèles neuronaux profonds (qu'ils soient à propagation avant, convolutionnels ou récurrents) ne sont pas totalement compris d'un point de vue théorique (impact du nombre de neurones et de couches sur l'approximation produite, nature de la dépendance temporelle des données traitées). Un des axes que nous souhaitons explorer consiste à aborder l'étude de ces architectures neuronales au travers des *liens* qu'elles présentent avec certaines modélisations probabilistes markoviennes, l'objectif étant d'englober ces deux modélisations dans un cadre commun dans l'espoir, d'une part, de construire des modélisations markoviennes plus puissantes, et d'autre part de mieux comprendre le résultat fourni par les architectures neuronales.

Un second axe consiste à *mieux prendre en compte les spécificités structurelles des images* au sein même du réseau de neurones. Nous pensons pour ce faire modifier la structure des réseaux de neurones convolutifs, les filtres linéaires pouvant le cas échéant être remplacés par des outils de morphologie mathématique (en particulier des opérateurs connexes), ce qui, en retour, posera la question de l'apprentissage de telles structures.

Les méthodes de classification par apprentissage profond reposent le plus souvent sur l'abondance des jeux d'entraînement et la qualité de leurs annotations. Néanmoins, les apprentissages conduits sur des jeux de données trop restreints ou de mauvaise qualité risquent de conduire à des problèmes de sur-apprentissage. Ceci représente un vrai verrou technologique, en particulier pour les applications (telle la détection d'objets d'intérêt dans des contenus multimédias) qui nécessitent de prendre en compte des ensembles évolutifs et personnalisables de catégories. Nous pensons aborder ce problème d'*apprentissage semi-supervisé* par des techniques d'apprentissage actif, ce qui posera à son tour le problème de la stratégie de sélection des données à étiqueter.

#### Hybridation apprentissages symbolique / statistique

L'Intelligence artificielle (IA) symbolique consiste à doter un système de mécanismes de raisonnement capables de manipuler les données symboliques représentant les connaissances et l'expertise d'un domaine. Cette approche fait appel aux modèles et méthodes de logique. L'apprentissage en IA symbolique se fait directement sur des règles connues contrairement à l'apprentissage statistique qui se base sur l'observation et l'expérimentation du monde pour en déduire des règles. Ces deux visions sont complémentaires. Le défi majeur auquel nous nous proposons de contribuer est le développement de nouvelles approches hybrides permettant la production de modèles d'IA plus explicables et interprétables par l'humain en particulier pour le traitement de grandes masses de données dynamiques et évolutives (flux de données). Il s'agit d'une part, pour les algorithmes d'apprentissage de s'adapter

en permanence aux conditions environnementales changeantes (y compris leurs propres conditions) et aux besoins changeants des utilisateurs. Et d'autre part, le processus d'inférence d'information, qui transforme les flux de données en faits tangibles et les combine à des connaissances de base pour déduire de nouvelles connaissances, doit être effectué de manière continue et en ligne. Les solutions existantes souffrent de besoins en ressources mémoire et processeurs qui croissent non linéairement avec la taille des données susceptible d'être infinie dans un flux. Explorer plus finement cette fertilisation croisée entre les deux paradigmes permettra d'améliorer et renforcer la synergie entre les différentes équipes du laboratoire.

### **Apprentissage statistique et compression**

Nous pensons également utiliser l'apprentissage statistique pour améliorer certaines techniques de codage (codage d'images à très bas débit, codage de nuages de points 3D) et, réciproquement, proposer des algorithmes de compression de données issues de l'apprentissage statistique. Plus précisément :

- L'utilisation de réseaux 2G-SMS, encore largement répandus dans les pays émergents, pose la question de la compression très bas débit d'images fixes. Les pistes identifiées sont l'exploitation des réseaux de neurones profonds pour optimiser la chaîne de codage, et la mise en oeuvre de nouveaux paradigmes de compression fondés sur des réseaux de type GAN (Generative Adversarial Networks) ;
- Les avancées récentes des technologies de reconstruction 3D à partir d'images ou d'acquisitions laser ont conduit à une utilisation accrue de nuages de points 3D. Nous souhaitons développer des algorithmes d'apprentissage statistique en vue de produire des codecs comportant un contrôle fin du nombre de points 3D reconstruits ;
- L'essor des approches basées sur l'apprentissage profond et le besoin de les utiliser dans un contexte embarqué ou mobile conduit à la problématique nouvelle de la compression des paramètres d'un tel réseau. Nous nous intéresserons à l'adaptation des techniques de compression de signal pour ce nouveau type de contenu et leur transmission temps réel.

### **Apprentissage statistique pour la prédiction et la détection en vue d'usages réseaux et sécurité**

Nous souhaitons explorer l'adéquation et l'adaptation d'algorithmes d'apprentissage profond ou par renforcement pour faire avancer l'état de l'art sur certains problèmes d'infrastructures de réseaux, de services et fonctions réseaux virtualisés et de sécurité. À partir de sondes disséminées sur les infrastructures physiques et/ou virtuelles, nous comptons collecter de gros volumes de données (souvent non structurées) reflétant à la fois le trafic réseau et l'état des ressources. L'idée est ainsi de nous appuyer sur l'analyse statistique de ces données et la caractérisation des données de monitoring pour prédire la charge de trafic, détecter les anomalies et les écarts au SLA (Service Level Agreement) pour les réseaux 5G, et aussi pour détecter des anomalies liées à des comportements malveillants de manière à classifier les entités de façon fiable et efficace. L'objectif de ces travaux est d'optimiser la boucle de monitoring, d'analyse, de décision, de planification et enfin d'exécution d'actions correctives ou d'adaptation dynamique des infrastructures.

### **Traitement de données massives dans des grands graphes**

Le problème de dynamique d'opinion continuera à mobiliser quelques équipes du laboratoire. Selon les applications, l'opinion peut par exemple représenter la confiance par rapport à un agent/opérateur, ou une note d'un produit de consommation, ou un avis politique, ou une valeur mesurée par un capteur, etc. Ces opinions se propagent à travers le graphe/réseau selon plusieurs modèles possibles. Pour les évaluer au mieux, il convient d'appréhender leur mode de propagation en utilisant les données existantes. Trouver une représentation parcimonieuse des opinions dans un grand graphe en sélectionnant par exemple un sous-ensemble de noeuds/agents représentatifs est un autre verrou qu'on pourrait relever.

### **2.3.3 Projet scientifique de l'action spécifique "Internet des objets"**

Des études de marché récentes estiment jusqu'à 50 millions le nombre d'objets connectés d'ici 2020. Les réseaux de communication futurs devront par conséquent intégrer les contraintes spécifiques à l'IdO pour que ces prévisions deviennent une réalité. Dans cette AS, le problème va être étudié notamment sous les aspects liés aux réseaux, aux intergiciels et à la sécurité.

## Aspects réseaux

- *Optimisation de la consommation d'énergie dans les réseaux mobiles et IdO*. Une approche permettant de gérer l'énergie de façon parcimonieuse, mais également d'assurer un fonctionnement en cas de perte ou de non accès aux infrastructures, est d'autoriser la communication directe entre objets. Une première activité envisagée est donc la transmission coopérative pour les communications d'objet-à-objet (device-to-device), qui consiste à exploiter au mieux la réutilisation spatiale du spectre sans passer par une station ou une antenne relai, excepté en bout de chaîne ; des algorithmes spécifiques de traitement du signal ou d'optimisation pourront être développés à cet effet. Une deuxième activité, qui vient renforcer la première, est l'optimisation des systèmes radio sur le plan de l'efficacité de propagation et de la réduction de la consommation. Au niveau des systèmes électroniques, nous privilégierons les approches radio logicielles (SDR) pour leur capacité à traiter le contexte multistandard / multibande auquel nous sommes confrontés.
- L'AS IdO traitera aussi du *nommage des objets*. Les approches "Content Centric Networks (CCN)", "Information Centric Networks (ICN)" s'appuyant sur un nommage sémantique des objets et des systèmes de résolution de noms ne sont pas adoptés par les acteurs et n'ont pas intégré les infrastructures opérationnelles ou de production. Il en est de même pour l'initiative "Location and Identity Separation Protocolz (LISP) qui a consisté à séparer l'identité de la localisation physique des objets (c'est à dire, la manière d'y accéder physiquement par le réseau). Ce protocole se fonde sur le principe du Domain Name System (DNS) qui a le grand avantage de passer à l'échelle. LISP, pour gérer la mobilité des objets, introduit une complexité et une charge lourde sur les infrastructures. Il convient de revoir nos approches et c'est l'essence même des travaux qui sont envisagés au sein du laboratoire, avec comme socle la structure du DNS et la gestion des identités et des accès (commande et contrôle). Nous comptons adapter le DNS à l'IdO en traitant de manière conjointe nommage et identité, tout en assurant la séparation et en minimisant la complexité.
- Les réseaux mobiles de cinquième génération (5G) sont pressentis pour le *transport de nouveaux services*, dits *critiques*, tel que le MTC (machine type communication), qui inclut l'IdO et Industry 4.0 par exemple, et implique des milliards d'objets avec des contraintes fortes sur la latence, la fiabilité, et l'énergie. Nous étudierons dans ce contexte les mécanismes de transport efficaces et scalables, ainsi que le partage de ressources entre les services dans le cadre du concept de slicing.

## Aspects Intergiciels

Un deuxième axe de l'AS IdO est le développement d'**intergiciels pour l'IdO** qui adresse le problème du nombre d'objets connectés et leur distribution à différentes échelles (nuage, géorépartition, etc.) et du volume de données à filtrer et traiter dans un temps contraint. Ce qui pose de nombreux défis en terme d'intergiciels et de génie logiciel.

Dans ce contexte, un premier axe de travail porte sur la conception des services de découverte à très grande échelle pour la sélection d'objets et de données. Ces services permettront la sélection des données ayant un niveau de qualité satisfaisant tout en respectant la vie privée des utilisateurs. D'un point de vue génie logiciel, nous envisageons le couplage de ces services de sélection avec des intergiciels côté client qui s'interfacent de manière transparente avec la plateforme IdO ou le service de découverte. L'objectif est d'offrir le bon niveau d'abstraction côté applicatif tout en assurant la gestion des interactions avec les services de découverte de manière transparente et économique du point de vue de l'utilisation de la batterie.

Un deuxième axe concerne le découplage des systèmes répartis construits avec le paradigme de l'IdO : découplage en termes de temps, d'espace et de synchronisation, des consommateurs, des producteurs et des unités de traitement. Ces caractéristiques accroissent les problèmes de nommage, de découverte, de filtrage, de routage, de déploiement à la demande, de tolérance aux fautes et aux déconnexions, et de passage à l'échelle. Nos travaux ont commencé à aborder ce défi par la différenciation de la dissémination des informations, du placement des traitements selon les dimensions et les échelles considérées. Les premières dimensions et échelles prises en compte concernent les objets connectés avec faible capacité de communication et de traitement ; passerelles et serveurs de proximité géorépartis ; serveurs de calcul en nuage ; dispositifs des utilisateurs. Nous prévoyons de continuer l'étude des systèmes multi-échelles en ajoutant la prise en compte de la forte dynamique due à la nomadicité des clients ainsi qu'à la mobilité des objets et des unités de traitement.

## Sécurité et protection des données personnelles

Un troisième axe de l'AS IdO est la **sécurité** et la **protection des données personnelles** qui adresse le problème de la prise de contrôle à distance des objets, de la réalisation d'attaques logiques vers ou depuis les objets, de la perturbation du fonctionnement attendu du système, et de la fuite de données personnelles ou industrielles, tous ces phénomènes ayant un impact plus ou moins critique sur les entreprises, les États et les individus suivant les domaines d'applications considérés : "Industrie du futur, Automobile, Santé". Le grand nombre d'objets, leur hétérogénéité, leur accessibilité physique et les contraintes en ressources nécessitent de concevoir de nouvelles approches ou d'adapter en profondeur les solutions traditionnelles.

Concernant la *sécurité*, plusieurs directions sont envisagées. Pour la détection d'attaques, il s'agit de distinguer une anomalie d'une attaque grâce à des techniques d'apprentissage profond, ceci pour ajuster la réponse à l'incident, et aussi de prendre en compte l'hétérogénéité des objets dans les opérations de surveillance et de réaction. Une fois l'attaquant détecté, il s'agit de l'isoler en s'appuyant sur des approches de gestion de la confiance ou sur la programmabilité des réseaux SDN. En cas d'attaques réussies, il est proposé de définir un mode de fonctionnement dégradé du service à la condition que la sécurité du système ne soit pas remise en cause.

Concernant la *protection des données personnelles*, nous prévoyons plusieurs axes de travail. Pour répondre aux besoins de se conformer au RGPD, un premier axe porte sur la gestion du consentement de l'utilisateur, sous les angles d'une preuve infalsifiable émise par l'utilisateur, et de la gestion des obligations attenantes dans un système publish/subscribe. Un autre axe concerne la personnalisation de services qui ne doit s'appuyer que sur la technologie (et non le droit) pour préserver la vie privée. Enfin, pour pallier au problème d'hétérogénéité, nous proposons un travail d'alignement des ontologies de référence et des mécanismes d'analyse sémantique à exécuter dans l'objet.

# 3 Équipe ACMES

ACMES (Algorithmes, Composants, Modèles Et Services pour l’informatique répartie) est une équipe regroupant les enseignants-chercheurs en informatique du laboratoire SAMOVAR. Elle mène diverses recherches en informatique sur les sujets suivants : les infrastructures logicielles (acquisition, stockage, middleware, performances), la gestion des processus métiers, la gestion des données, la modélisation des raisonnements et l’extraction des connaissances (fouille, apprentissage). Elle est organisée en quatre axes qui sont en interaction : Intergiciels et algorithmique pour les environnements ubiquitaires, intelligence collective et sociale, cloud et processus métier, et enfin parallélisme et système.

## 3.1 Présentation de l’équipe ACMES

### 3.1.1 Introduction

L’équipe ACMES s’intéresse principalement à la définition de modèles, d’algorithmes et d’outils pour la gestion des données et des connaissances, et à la construction d’applications réparties sur des architectures à grande échelle. Les solutions visées doivent tenir compte de plusieurs propriétés : du volume massif, hétérogène et dynamique des données, du caractère diffus des environnements d’exploitation, de la diversité des services et architectures logicielles, mais aussi de l’hétérogénéité des technologies d’accès et de communication. Au regard de la complémentarité des compétences des quatre groupes de l’équipe ACMES, notre démarche aborde cette problématique dans un continuum allant de la couche application jusqu’à la couche réseau en passant par les couches middleware et système.

Plus précisément, les compétences de l’équipe ACMES dans ce quinquennat s’inscrivent dans les thématiques suivantes :

- Parallélisme et système : Calcul haute performance, multicoeur, virtualisation, système d’exploitation ;
- Les systèmes répartis : Algorithmes parallèles et répartis, intergiciel, agents mobiles, services computing ;
- La gestion des procédés métiers : Cycle de vie des applications, portabilité et reconfiguration des applications, gestion de l’allocation de ressources, élasticité du Cloud ;
- La gestion des données et des connaissances : Traitement de données (en flux, en graphes, etc.), représentation et extraction de connaissances, modélisation des raisonnements, gestion de conflit et de l’incertitude.

Nos applications privilégiées sont l’habitat intelligent, la ville intelligente, l’aide aux personnes, l’industrie du futur et la santé.

### 3.1.2 Effectifs et moyens de l’équipe

Depuis le précédent quinquennat, l’équipe s’est renforcée avec l’intégration de quatre maîtres de conférences (Elisabeth Brunet le 01/10/2015, Amina Guermouche le 01/10/2016, Christian Parrot le 01/01/2015, François Trahay le 01/10/2015) et un professeur (Gaël Thomas le 01/01/2015) sur le thème parallélisme et système dont l’objectif de recherche est de rendre l’utilisation massive de ressources informatiques à la fois plus performante et plus facilement programmable. Cet axe vient consolider les thématiques de l’équipe avec des techniques de traitement parallèle de données massives et d’élaboration de techniques d’optimisation à un niveau global. De plus, l’équipe a recruté un maître de conférences (Badran Raddaoui le 01/10/2017) sur la thématique intelligence artificielle pour renforcer l’axe intelligence collective et sociale et un maître de conférences (Pierre Sutra le 01/10/2015) sur l’axe

intergiciels et algorithmique pour les environnements ubiquitaires.

Par ailleurs, l'équipe a vu aussi le départ de l'un de ses membres (Samir Tata qui a rejoint IBM Almaden Research Center en septembre 2016) tout en restant membre associé depuis.

Finalement, l'effectif actuel de l'équipe ACMES est constitué comme suit : 17 membres permanents (8 professeurs et directeurs d'études et 9 maîtres de conférences) et 2 chercheurs associés, avec 47 doctorants et 15 post-doctorants sur la période de référence.

**Professeurs et directeurs d'études IMT :** Djamel Belaïd, Amel Bouzeghoub (HDR), Sophie Chabridon (HDR), Bruno Defude (HDR), Walid Gaaloul (HDR), Claire Lecocq, Samir Tata (HDR, membre jusqu'à septembre 2016), Gaël Thomas (HDR), Daqing Zhang.

**Maîtres de conférences :** Elisabeth Brunet, Denis Conan (HDR), Alda Gańczarski, Amina Guermouche, Christian Parrot, Badran Raddaoui, Pierre Sutra, Chantal Taconet (HDR), François Trahay.

**Chercheurs associés :** Samir Tata (depuis septembre 2016), Sadok Ben Yahia (HDR, FST Tunis, depuis janvier 2012). Le bilan ne tient compte que de leurs publications co-signées avec un membre de l'équipe.

Le tableau suivant présente l'évolution du personnel enseignant-chercheur de l'équipe ACMES sur la période 2013-2018

	1/1/2013	1/1/2014	1/1/2015	1/1/2016	1/1/2017	30/6/2018
Enseignants chercheurs permanents	11	11	16	16	17	17
CDD ingénieurs (en mois)	24	3	26	10	35	0
Chercheurs associés	0,5	0,5	0,5	0,5	1	1
Doctorants	23	19	19	18	20	19,5
Post-docs (en mois)	43	19	36	12	1	0

Le nombre de doctorants se stabilise autour d'une vingtaine par an et le nombre de post-doctorants présents annuellement dans l'équipe est de l'ordre de 3 à 4 selon les années avec des périodes de présence d'une moyenne de 9 mois. À ces jeunes chercheurs, il convient d'ajouter l'appui d'ingénieurs de recherche recrutés en CDD dont la durée dépend des types et des volumes de contrats.

Le tableau suivant donne le nombre de thèses soutenues et en cours sur la période de référence ainsi que le nombre de HDR soutenues. Il y a en général entre 4 et 5 thèses soutenues par an (ce nombre a doublé en 2015) et une quinzaine de thèses en cours par an. Il est à noter que l'équipe est montée en puissance avec 3 HDR supplémentaires.

	2013	2014	2015	2016	2017	2018	Total
Thèses soutenues	4	4	9	4	5	4	30
Thèse en cours	19	15	10	14	15	15,5	-
HDR soutenues	0	2	1	0	0	0	3

La durée moyenne des thèses est de 41 mois sur la période 2013-2018. Les mois d'extension sont inhérents aux types de recherches réalisées dans l'équipe qui nécessitent souvent des réalisations logicielles et des expérimentations longues et indispensables à la validation des travaux développés.

Le mode de financement des thèses en cours est décrit dans le tableau suivant :

	2013	2014	2015	2016	2017	2018	Total
IMT-TSP	7	3	1	0	1	0	12
co-tutelle (E/C étrangers)	1	1	2	2	2	1	9
Gouvernement étrangers	5	0	0	0	0	0	5
Externe	1	1	1	0	0	0	3
CIFRE	2	0	1	0	2	0	5
Contrats	7	0	0	1	1	1	10
Organismes (CEA)	1	0	0	2	0	0	3

Ce financement se répartit ainsi : environ 2/3 provient des bourses de l'institut Mines-Telecom (ressourcement Carnot, programme Futur et Rupture, bourses TSP) et des bourses externes (gouvernements étrangers, cotutelles, Erasmus). Le tiers restant est un financement industriel (CIFRE, contrats industriels, CEA). Peu de doctorants sont financés exclusivement sur contrats; ces derniers servant souvent à allouer un complément aux boursiers de gouvernements étrangers dont les montants sont souvent inférieurs à ceux des allocations françaises.



Les autres ressources financières sont données dans le tableau suivant (en K€). L'équipe a une capacité de financement sur ressources externes qui s'élève en moyenne à 400 K€/an avec une équi-répartition entre ANR, pôles de compétitivité (et investissement d'avenir), contrats industriels et l'Europe.

	2013	2014	2015	2016	2017	Total
Projets ANR	64	78	160	6	16	325
Poles de compétitivité	0	216	23	274	2	515
Projets européens (ITEA, FP7...)	374	141	13	0	0	528
Autres financements publics sur appels à projets	102	91	90	105	137	525
Contrats industriels	95	61	60	137	85	438

### 3.1.3 Profils d'activité

- Activité scientifique : 40% en moyenne,
- Activité de valorisation et de transfert : 25% en moyenne
- Participation active à la formation par la recherche : 20% en moyenne
- Appui à la communauté : 15% en moyenne

Il faut noter que l'implication dans l'une ou l'autre des activités est variable selon les personnes et les années, en fonction notamment des charges d'enseignement qui sont modulables d'une année sur l'autre.

## 3.2 Produits de la recherche et activités de la recherche de l'équipe ACMES

D'un point de vue quantitatif, l'ensemble des publications scientifiques de l'équipe s'élève à 360 publications en conférences et revues, ce qui, pour 17 enseignants-chercheurs permanents sur une période d'évaluation de 5 ans, donne un taux de production d'environ 4 par membre et par an. Un effort particulier de publication en revues a été fait (36% pour ce quinquennat contre 25% pour le précédent) et devrait s'accroître avec l'étoffement de l'équipe. Notons également que la qualité des publications s'est maintenue à un bon niveau pour les conférences qui constituent un lieu de publication privilégié dans nos domaines de recherche (64%).

Le tableau suivant présente la répartition des publications sur la période 2013-2018.

	2013	2014	2015	2016	2017	2018	Total
Revues classées A* ou A	9	5	7	11	5	3	40
Revues classées B	5	10	9	8	8	5	45
Autres revues	4	8	7	10	11	3	43
Conférences classées A* ou A	14	9	15	29	22	11	100
Conférences classées B	15	5	9	9	13	4	55
Autres conférences/colloques	19	18	13	12	14	1	77
Chapitres de livres	3	2	1	1	0	1	8

D'un point de vue qualitatif, la stratégie de publications a consisté à viser prioritairement les supports de publication les plus prestigieux du domaine au niveau international, en nous basant sur le classement ERA CORE 2018 pour les conférences et ERA CORE 2010 pour les revues. Ainsi, près d'1/3 des revues et 43 % des conférences sont classés A ou A\*. Certaines conférences ont un taux de sélectivité et un facteur d'impact très élevés (les conférences A\* et certaines conférences A).

La publication de nos travaux de recherche dans les colloques et autres conférences nationales non sélectifs demeure néanmoins très importante pour la présence dans la communauté, l'animation ou la formation des doctorants.

### 3.2.1 Bilan scientifique

Nous articulons la présentation des travaux réalisés au sein de l'équipe autour des quatre thématiques principales : Parallélisme et Système, Intergiciels et algorithmique pour les environnements ubiquitaires, Cloud et processus métier, Intelligence collective et sociale. Pour respecter les contraintes imposées par l'Hcéres, nous n'avons fait figurer qu'une sélection de 20% des publications.

## Parallélisme et Système

**Membres :** Elisabeth Brunet, Denis Conan, Amina Guermouche, Christian Parrot, Pierre Sutra, Gaël Thomas, François Trahay ;

Au cours de la dernière décennie, les architectures informatiques ont profondément évolué. Les serveurs sont devenus multicœurs et disposent souvent d'accélérateurs matériels contenant plusieurs centaines d'unités de calcul et permettant d'exécuter des codes massivement parallèles (GPU, Xeon Phi). Dans le domaine du parallélisme et de la distribution des systèmes informatiques, les grands défis scientifiques concernent le passage à l'échelle avec le nombre d'unités de calcul, la facilité de programmation de ces architectures et la compréhension des performances. Ces dernières années, le groupe parallélisme et système s'est intéressé à ces défis en proposant de nouveaux outils, langages et modèles, ainsi que des algorithmes afin de mieux comprendre et exploiter ces nouvelles architectures.

**Analyse de performance des architectures multicœurs** L'analyse de performance des architectures multicœurs a été étudiée sous divers aspects. Dans [32.C.ac], nous avons introduit une approche d'analyse des effets du chevauchement des calculs avec une communication point à point non bloquante. Dans [17.J.ac], un mécanisme de migration dynamique de processus est défini afin de réduire le temps d'échange et de transfert de données entre les différents composants d'une application. Ce mécanisme est basé sur un pattern d'accès bloc similaire à l'utilisation d'un fichier local d'échange de données. L'idée étant de faire migrer les processus utilisant le même pattern de communication vers le même nœud pour que leurs échanges de données se fassent via des fichiers locaux et non globaux. Un algorithme permettant une telle migration a été proposé. Dans le cadre des environnements Cloud, une technique de lecture anticipée de données sur les serveurs de stockage a été introduite dans [3.J.ac]. Deux algorithmes de prédiction ont été proposés à cet effet pour déterminer, par anticipation, les blocs de données qui doivent être récupérés sur les serveurs de stockage. Les expérimentations ont montré qu'un tel mécanisme pouvait contribuer efficacement à la performance du système. Dans [41.C.ac], un outil d'analyse des fichiers de traces d'évènements a été développé afin de mesurer les performances des architectures multicœurs. Cet outil implémente un algorithme de détection de pattern évènements qui permet de représenter une trace comme une boucle sur un ensemble d'évènements. Cet outil a été intégré au framework d'analyse de performance EZTrace.

Enfin, dans le cadre de la thèse de M. Said Mosli Bouksiaa, un outil orienté-effet a été développé afin de quantifier l'impact des interférences dans les applications multi-thread. En effet, des interférences peuvent survenir dans ce type d'application en raison des accès multiples à la même ressource ralentissant ainsi l'exécution des threads. L'outil proposé vient en complément des outils de profiling existants qui se focalisent sur l'identification des causes de l'interférence, et non pas sur ses effets. Le développeur ne peut donc pas conclure si l'optimisation d'une interférence identifiée peut produire une meilleure performance. L'évaluation de l'outil sur 27 applications a permis d'identifier 12 goulets d'étranglements (bottlenecks) causés par 6 types d'interférence différents.

Certains de ces travaux ont été menés en collaboration avec Southwest University (Chongqing, Chine) et l'Université de Tokyo (Japon).

**Optimisation des codes systèmes pour les architectures multicœurs** Dans ce contexte, le groupe parallélisme s'est intéressé à optimiser les algorithmes de verrouillages en accélérant les sections critiques des applications basées sur les architectures multicœurs [16.J.ac]. L'idée est de remplacer l'acquisition d'un verrou par un appel de procédure à distance sur un cœur dédié. L'algorithme proposé, appelé Remote Core Locking (RCL), limite l'effondrement des performances observé avec d'autres algorithmes de verrouillage lorsque de nombreux threads tentent d'acquérir un verrou en même temps, et supprime le besoin de transférer des données protégées par un verrou vers le cœur acquérant le verrou. D'autres travaux ont consisté à étudier le passage à l'échelle de mémoires transactionnelles logicielles (STM) [1.C.ac].

Le groupe a également contribué à l'optimisation de la gestion mémoire dans les machines virtuelles Java en proposant de modifier la machine virtuelle JAVA pour éliminer les références inutilisées dans les applications dynamiques OSGi [43.C.ac]. L'équipe a aussi travaillé sur l'optimisation des pointeurs intelligents C++ (smart pointers) en utilisant les mémoires transactionnelles [35.C.ac]. Enfin, l'équipe a contribué à l'optimisation des algorithmes de ramasse-miettes (*garbage collector*) afin d'améliorer leur débit sur les architectures multicœurs [42.C.ac, 19.C.ac, 40.C.ac].

Dans le cadre de la thèse de Gauthier Voron [16.C.ac], une approche a été définie afin de rendre la couche de virtualisation utilisée dans l'informatique dématérialisée plus efficace pour les machines multi-cœurs reposant sur une architecture Non Uniform Memory Access (NUMA). En effet pour ce type d'architecture, le placement des

tâches et des données sur les cœurs est déterminant pour les performances. Le mauvais placement actuellement effectué par les couches de virtualisation grève les performances et peut multiplier les temps de calcul par 7. L’approche proposée consiste à définir des politiques de placement efficaces dans l’hyperviseur Xen et les résultats expérimentaux montrent que cette solution élimine la dégradation due à la virtualisation sur les architectures NUMA.

Certains de ces travaux ont été menés en collaboration avec l’Université de Neuchâtel (Suisse), avec Oracle Labs et avec l’Université de Knoxville (USA).

**Programmabilité et modélisation des architectures multicœurs** Le groupe parallélisme a proposé de nouveaux langages permettant de simplifier la programmation des accélérateurs matériels en développant l’outil STEPOCL et son langage associé qui permettent de simplifier le développement d’applications pour des accélérateurs multiples [39.C.ac]. Le groupe a également contribué à la définition du modèle de programmation dSTEP, basé sur les directives, pour les machines à mémoires partagées, distribuées et hybrides afin d’améliorer la productivité du programmeur tout en garantissant de bonnes performances en termes de temps d’exécution et de l’usage de la mémoire [14.J.ac]. Ces travaux ont fait l’objet des thèses de Pei Li et Rachid Habel.

Dans [9.C.ac], le système CRESON, pour l’invocation d’objets stockés dans des bases de données NoSQL a été proposé dans le but de réduire le coût du mapping Objet-NoSQL. La cohérence entre l’objet et sa copie est assurée à l’aide du concept de *state machine replication*. Un tel système est particulièrement intéressant dans les environnements Cloud où le stockage persistant d’objets favorise le développement d’applications.

Des travaux ont été aussi menés sur les briques de base composant ces systèmes (e.g., primitives de communication, détecteurs de défaillances) afin de proposer de nouvelles abstractions [26.J.ac], d’en étudier la complexité [24.C.ac, 7.J.ac] et l’applicabilité [7.C.ac].

Ces avancées ont été possibles grâce à plusieurs collaborations : RedHat, l’Université de Rennes, l’Université Rovira y Virgili (Espagne), l’Université Catholique de Louvain (Belgique) et l’Université d’Uberlandia (Brésil).

## Intergiciels et algorithmique pour les environnements ubiquitaires

**Membres** : Djamel Belaïd, Sophie Chabridon, Denis Conan, Claire Lecocq, Pierre Sutra, Chantal Taconet ;

Nous nous intéressons aux algorithmes et services intergiciels pour les environnements ubiquitaires. Sur la période, nous nous sommes intéressés particulièrement à trois thématiques complémentaires concernant l’Internet des objets : la distribution multi-échelle du contexte des objets de l’IoT, la qualité du contexte et la protection de la vie privée ainsi que l’ingénierie dirigée par les modèles des applications IoT.

### Distribution multi-échelle des informations de contexte en provenance de l’Internet des Objets

Dans l’IoT, le nombre d’objets connectés croît de manière exponentielle et les objets sont très volatiles et mobiles. Ces objets peuvent être des sources (ou producteurs) d’informations sur leur contexte et fournir des informations à transformer et à acheminer vers un nombre important et fluctuant de consommateurs. Ainsi, la gestion des informations de contexte intervient à de multiples échelles (réseau personnel, Internet, nuages) avec un découplage fort entre producteurs et consommateurs.

La très grande quantité de données, fournie par les producteurs de contexte d’une manière découplée et asynchrone, exige des techniques de filtrage tenant compte des exigences sur la vie privée des personnes concernées et capables de fournir le niveau de QoC (qualité de contexte) attendu par les applications [25.C.ac].

Par ailleurs, comme les systèmes sont fortement répartis, par différence des solutions basées sur l’informatique en nuage, l’approche pour gérer de nombreux producteurs et consommateurs ainsi qu’une grande quantité de données est le « passage à l’échelle localisé » introduite par Satyanarayanan<sup>1</sup>. Le concept d’échelle (cf. les contributions en ingénierie dirigée par les modèles présentées ci-dessous) est projeté sur le concept de portée de distribution (en anglais, *scope*) des systèmes répartis à base d’événements : une portée de distribution est une abstraction qui rassemble des clients (producteurs et consommateurs) dans un ensemble tel que les notifications des producteurs sont visibles des consommateurs de la même portée ; les portées de distribution sont organisées en hiérarchie avec du filtrage pour traverser les frontières des portées<sup>2</sup>. Ainsi, les informations de contexte qui ne sont pas « demandées »

1. M. Satyanarayanan, “Pervasive Computing : Vision and Challenges”, IEEE Personal Communications, volume 8, number 4, pages 10–17, August 2001.

2. L. Fiege, M. Cilia, and B. Mühl, “Publish-subscribe grows up : Support for management, visibility control, and heterogeneity”, IEEE Internet Computing, volume 10, number 1, pages 48–55, January 2006.

dans une partie du système ne donnent pas lieu à la diffusion de souscription ou à la transmission de notifications dans les parties du système réparti ne contenant pas de consommateurs intéressés [17.C.ac]. L'apport de nos travaux est double : lier la modélisation des portées de distribution dans plusieurs dimensions avec l'ingénierie dirigée par les modèles des applications en cours de construction, et organiser algorithmiquement la répartition des producteurs et des consommateurs dans de plusieurs graphes de portées de distribution.

Nos contributions sont évaluées et validées à travers les canevas logiciels MUDEBS<sup>3</sup> et MUCONTEXT<sup>4</sup>, qui constituent aussi des briques logicielles de la plateforme M4IoT (*Middleware for the Internet of Things*)<sup>5</sup>.

**Qualité de contexte et protection de la vie privée dans l'Internet des objets** Le déploiement de l'informatique sensible au contexte dans les systèmes ambiants qui nous entourent rend possibles de nouvelles applications. Ainsi apparaît une intelligence ambiante où les objets de la vie courante deviennent capables de déclencher une action selon le contexte, sans interaction avec l'utilisateur. Les informations de contexte étant ainsi au cœur de l'intelligence ambiante pour prendre des décisions d'adaptation, il devient primordial de définir des mécanismes pour limiter l'impact que peuvent avoir leurs imperfections intrinsèques liées aux limitations des capteurs. Une solution est d'ajouter des méta-données de qualité aux informations de contexte et ensuite de considérer la qualité de contexte (Quality of Context ou QoC) en même temps que le contexte lui-même lors du traitement et de l'analyse du contexte.

L'Internet des objets (IoT) donne une nouvelle dimension à l'intelligence ambiante en prenant en compte également l'environnement éloigné accessible par un réseau longue distance.

Le cadre QoCIM a été proposé comme une solution de gestion de la QoC à la fois générique, calculable et expressive pouvant s'intégrer au sein de gestionnaires de contexte répartis pour l'IoT [50.C.ac], [47.C.ac]. L'évaluation de la QoC faisant appel à des ressources de calcul limitées, des solutions flexibles et efficaces doivent être mises en place pour manipuler la très grande quantité de données et calculer uniquement les critères de QoC pertinents à un moment donné. Une approche autonome a été proposée et évaluée pour reconfigurer dynamiquement la gestion de la qualité de contexte en fonction des ressources disponibles [28.C.ac].

Afin de réaliser une gestion de contexte à l'échelle de l'IoT, les producteurs et les consommateurs de contexte doivent être découplés à la fois en temps, en travaillant à des moments et des vitesses différents, et en espace, en n'ayant pas besoin de se connaître de manière directe. Ceci nécessite de nouvelles solutions pour la protection de la vie privée des personnes sur lesquelles portent les données de contexte [24.J.ac].

Nous avons proposé des modèles de contrats de contexte permettant de traiter conjointement la QoC et le respect de la vie privée du point de vue des producteurs et des consommateurs de contexte. Dans une approche d'ingénierie dirigée par les modèles, ces contrats permettent ensuite de générer des filtres de routage, déployés au sein d'un gestionnaire d'événements distribués, pour limiter les flux d'événements diffusés.

**Ingénierie dirigée par les modèles pour construire des applications dans le contexte de l'Internet des objets** Un processus de conception, basé sur l'ingénierie dirigée par les modèles (IDM), a été défini afin de générer automatiquement les éléments logiciels requis par la gestion de la qualité de contexte [27.J.ac].

Dans le cadre de l'Internet des objets, les systèmes répartis considérés sont complexes et hétérogènes. Nous avons défini un processus de caractérisation multiéchelle pour qualifier des systèmes répartis suivant des points de vue tels que la dispersion géographique des entités, la nature des équipements qui les hébergent, les réseaux sur lesquels elles sont déployées, ou encore l'organisation des utilisateurs. Le processus conduit à un modèle de caractérisation multiéchelle et à la production d'artefacts logiciels qui apportent, à l'exécution, la conscience des échelles aux entités du système. Ces artefacts ont été utilisés notamment pour définir et gérer des contraintes multiéchelles lors du déploiement de systèmes complexes et pour structurer le routage d'informations de contexte au sein d'un système réparti multiéchelle [45.C.ac]. Le processus est outillé avec le cadre MUSCA<sup>6</sup>.

Ces travaux ont été réalisés dans le cadre de collaborations nationales et internationales notamment à travers le projet ANR INCOME et des thèses CIFRE : laboratoire IRIT (Université de Toulouse), laboratoire LII (ENAC),

3. <https://fusionforge.int-evry.fr/www/mudebs/>

4. <https://fusionforge.int-evry.fr/projects/mucontext/>

5. <https://www.telecom-sudparis.eu/recherche/plateformes-recherche/content/uploads/2017/04/M4IoT.pdf>

et <https://www.telecom-sudparis.eu/wp->

6. <https://fusionforge.int-evry.fr/www/musca/>

société Artal, EDF R&D Saclay, Université Fédérale du Rio Grande do Norte (UFRN Brésil), Université de Vigo (Espagne), Triagnosys/Safran (Münich, Allemagne).

## Cloud et processus métiers

**Membres :** Bruno Defude, Walid Gaaloul, Samir Tata (jusqu'à 09/2016) ;

Ce groupe de recherche s'intéresse principalement aux problématiques du développement et de la gestion d'applications distribuées à base de services, à la mise en place de mécanismes d'aide à la configuration correcte de processus métiers mais également au placement optimal de machines virtuelles (VMs), et ce dans un environnement Cloud.

**Développement et gestion d'applications distribuées à base de services** Nous avons principalement revisité les approches classiques de gestion du cycle de vie des applications distribuées à base de services pour les rendre compatibles avec le Cloud (principalement au niveau PaaS, Platform as a Service). Nos contributions ont porté sur la modélisation des applications, leur déploiement (portabilité, configuration) et leur exécution (allocation des ressources, placement, monitoring, reconfiguration, élasticité) en intégrant partiellement la partie données. Pour ce faire, nous avons adopté une démarche qui s'intègre dans l'approche SPD (Slice, Package, Deploy) proposée dans le cadre des travaux de thèse de Sami Yanguï [18.J.ac]. L'approche de slicing définie dans [44.C.ac] repose sur une modélisation abstraite de l'application, en termes de réseaux de Petri, qui est ensuite découpée en un ensemble de services élémentaires, gérés de manière autonome, tout en préservant la sémantique de l'application. Le packaging permet, d'une part, de minimiser l'empreinte de l'exécutable et, d'autre part, d'associer les services génériques nécessaires à l'exécution de l'application (monitoring, reconfiguration, élasticité notamment). Afin de permettre un déploiement qui favorise la portabilité vis-à-vis des fournisseurs de Cloud, nous avons proposé successivement une approche de broker [23.J.ac] puis une API et des wrappers basés sur le standard COAPS [11.J.ac], ainsi qu'une approche dirigée par les modèles pour la gestion de l'élasticité des ressources Cloud [4.C.ac, 11.C.ac].

Nous avons également contribué au monitoring et la reconfiguration d'applications basées sur le standard SCA dans le cadre de la thèse de Mohamed Mohamed en intégrant des services génériques additionnels permettant de gérer les services métiers, et s'appuyant sur le modèle OCCI que nous avons étendu au niveau plateforme et application [15.J.ac]. Dans le cadre de la thèse de Mourad Amziani, nous nous sommes intéressés à l'élasticité comme une des propriétés les plus importantes du Cloud. Nous avons proposé des mécanismes de gestion de l'élasticité au niveau procédé métier et non pas au niveau des VMs. L'intérêt majeur est de pouvoir gérer l'élasticité à un grain très fin puisque nous le faisons au niveau d'une activité. Plusieurs stratégies d'élasticité peuvent être supportées et nous avons proposé une formalisation à base de réseaux de Petri permettant de prouver la correction d'une application élastique vis-à-vis de certaines propriétés (pas d'introduction de deadlock ni de boucle d'élasticité notamment) [21.J.ac]. Ce travail sur l'élasticité a ensuite été étendu en proposant un modèle et un langage (STRATmodel et STRAT) de description de stratégies d'élasticité permettant leur comparaison et leur évaluation [21.C.ac].

Dans le cadre de la thèse de Rami Sellami, nous avons abordé le problème de l'interopérabilité de la gestion de données au niveau des plateformes PaaS. En effet, ces dernières années ont remis en cause le paradigme *one size fits all* des systèmes de gestion de bases de données avec l'utilisation, au sein d'une même application, d'un ensemble de gestionnaires de données spécialisés et donc hétérogènes complexifiant ainsi le processus de développement logiciel (systèmes relationnels et différentes catégories de systèmes NoSQL). Nous avons apporté plusieurs contributions à cette problématique : (i) définition d'une approche complète de déploiement dans le Cloud assurant principalement la portabilité des applications relativement aux systèmes de stockage utilisés et s'appuyant sur le modèle OCCI et l'API COAPS [13.J.ac], (ii) définition d'un modèle de données unifié et son algèbre associée, encapsulée par une interface de programmation de type REST, appelée ODBAPI, qui permet aux développeurs de s'abstraire de l'hétérogénéité des modèles de données et des langages de requêtes [13.J.ac], (iii) définition d'une approche d'intégration de données hétérogènes permettant l'expression et l'optimisation de requêtes multi-sources [2.J.ac]. L'ensemble de ces travaux ont été partiellement financé par les projets Monica, OpenPaaS et OCCIWare. Certains de ces travaux ont été menés en collaboration avec les universités de Paris 13 et d'Evry Val-d'Essonne, les universités de Monastir, Sfax et Tunis.

**Modélisation et gestion de processus métiers** Nos travaux s'intéressent également à la spécification d'applications à l'aide de processus métiers configurables avec comme objectif d'automatiser le soutien de la variabilité dans ce type de processus. Un modèle de processus configurable est un modèle générique qui intègre de

multiples variantes de procédés d'un même processus métier à travers des points de variation, appelés éléments configurables, qui permettent de multiples options de conception dans le modèle de processus.

Dans le cadre de la thèse de Nour Assy, nous avons contribué à comment assister (i) la conception des processus configurables d'une manière à ne pas perturber les concepteurs par des recommandations complexes [19.J.ac] et (ii) la création de systèmes de soutien de configuration afin de libérer les analystes de processus de la charge de les construire manuellement [46.C.ac]. Ce modèle de processus configurables a été étendu au contexte de l'IoT en définissant les applications utilisant des capteurs/actuateurs comme des processus métiers configurables, rendant ainsi la définition des applications indépendantes de leur déploiement et de leur exécution [20.J.ac]. Ce travail a fait l'objet de la thèse de Zahra Movahedi.

Par ailleurs, les entreprises adoptent de plus en plus des environnements Cloud pour déployer et exécuter leurs processus afin de bénéficier de ressources dynamiques à la demande. Dans un tel environnement multi-tenant, l'utilisation de modèles de processus configurables permet aux fournisseurs de processus Cloud de fournir un processus personnalisable. Dans ce contexte, nous avons proposé une approche pour supporter la modélisation et la configuration de l'allocation des ressources Cloud et des objets dans les modèles de processus configurables [31.C.ac, 2.C.ac]. Néanmoins, en l'absence d'une description explicite et formelle de la perspective ressources dans les processus métier existants, la correction de la gestion des ressources du Cloud ne peut pas être établie. Pour ce faire, nous avons adopté d'abord une approche basée sur les ontologies pour peupler sémantiquement une base de connaissances commune de processus et de ressources Cloud [8.J.ac, 8.J.ac]. Ensuite, nous avons élaboré une approche comportementale pour la vérification de la configuration tout en réduisant le problème bien connu de l'explosion d'espace d'états [6.J.ac]. Cette approche a permis d'extraire les options de configuration sans blocage d'un seul coup. Enfin, nous avons modélisé formellement l'allocation des ressources Cloud dans les modèles de processus métier [8.C.ac, 36.C.ac]. Cette modélisation a permis de vérifier la cohérence de l'allocation des ressources Cloud vis-à-vis des besoins des utilisateurs tout en prenant en compte les capacités disponibles des ressources. Finalement, nous avons adopté un algorithme génétique et une optimisation linéaire pour définir une allocation optimale des ressources Cloud aux processus métiers [26.C.ac, 12.C.ac]. Ces travaux ont fait l'objet des thèses de Emna Hachicha, Karn Yongsiriwit et Souha Boubaker.

Ces travaux ont été partiellement financés par Orange; certains ont été menés en collaboration avec le CEA, l'université de géo-sciences à Pékin, l'université de Vienne, l'université New South West Sydney, les universités de Monastir, Sfax et Tunis.

**Placement optimal de machines virtuelles** Nous nous sommes également attaqués au problème du placement optimal de VMs qui hébergent des applications hautement corrélées susceptibles d'être placées dans des centres de données (CDs) géographiquement distribués. Notre objectif était de prendre en compte les fluctuations de la demande impliquant des migrations de VMs tout en minimisant le volume du trafic global circulant entre les différents CDs. Nous avons proposé et validé, par simulation, des modèles d'optimisation stochastiques et déterministes pour traiter les différents modèles de trafic de communication. Des algorithmes quasi-optimaux qui permettent d'avoir la meilleure séquence de migration inter-CD des machines virtuelles inter-communicantes [4.J.ac] ont été ainsi mis en place. Ce sujet a fait l'objet de la thèse de Hana Teyeb.

## Intelligence Collective et Sociale

**Membres :** Amel Bouzeghoub, Alda Gancarski, Badran Raddaoui, Daqing Zhang, Sadok Ben Yahia (chercheur associé);

De nombreuses données sont généralement produites continuellement à partir de sources nombreuses (capteurs, sites web, objets connectés, etc.) et hétérogènes. Elles possèdent des caractéristiques qui complexifient leur traitement : structurées ou non, temporelles ou spatiales, statiques ou dynamiques (flux d'événements, évolution des réseaux sociaux), exactes ou incomplètes/imparfaites. Certaines de ces données sont porteuses de peu de sens si elles sont considérées isolément.

Le groupe Intelligence collective et sociale s'intéresse à la collecte, l'interprétation, la fusion et la corrélation de ces données brutes avec des données de contexte pour mieux les interpréter et produire les connaissances adéquates pour une meilleure prise de décision traduite par des actions du système. Cette prise de décision devient encore plus complexe lorsque l'environnement est incertain ou partiellement observé. La conception et la mise en œuvre de systèmes permettant cette transformation nécessitent l'intégration de compétences complémentaires dans les domaines de l'Intelligence Artificielle et de la Gestion de Données.

Nous présentons dans ce qui suit les principaux travaux du groupe organisés en quatre volets : le traitement des données en flux, le cycle de vie de la donnée et la prise de décision, l'analyse des réseaux sociaux et la collecte participative.

**Traitement des données en flux** Le processus de traitement des données en flux doit tenir compte de contraintes telles que l'espace de stockage limité par rapport aux données produites et un traitement en quasi temps réel. Les solutions visées doivent répondre aux exigences nombreuses de ces systèmes : réduction de la charge de données en entrée, traitement des données à la volée, stockage des données uniquement pertinentes pour l'analyse et l'extraction de connaissances utiles, etc. Les solutions existantes, comme les Systèmes de Gestion de Flux de Données (SGFD) ou les systèmes de traitement de flux de données sémantiques (RSP ou RDF Stream Processing), ne permettent pas de répondre à ces exigences. Notre approche de traitement de flux s'appuie également sur les RSP ; avec les contraintes suivantes : (1) garantir une certaine qualité de service lorsque les débits et/ou volumes de ces flux dépassent les capacités du système cible, (2) traiter le flux en une seule passe en utilisant exclusivement la mémoire vive et éviter tout swap vers le disque et (3) compresser/décompresser en ligne les flux de données RDF sans perte de données. Ceci a conduit à la définition de trois algorithmes complémentaires : un algorithme de détection de motifs fréquents (FreGraPad) pour extraire les séquences d'information pertinentes dans une fenêtre d'observation, un algorithme de délestage (loadshedding) basé sur les motifs produits et un algorithme de compression qui factorise les motifs fréquents. L'originalité du premier algorithme est sa structure de données résultant d'un codage des triplets RDF en une séquence binaire et qui permet une analyse du flux observé en une seule passe. L'originalité du second réside dans l'application d'opérations binaires pour élaguer les triplets non pertinents pour la requête continue. L'originalité du troisième est une compression sans perte de données qui peut être utilisée directement sans décompression dans l'évaluation de la requête. Appliquées en amont des moteurs de requêtes de type C-SPARQL ou CQUELS, ces techniques améliorent les performances de ces derniers (diminution de leurs temps de réponse, augmentation de leurs seuils de débit de traitement en entrée et optimisation de l'utilisation de leurs ressources systèmes [30.C.ac]). Ce travail a fait l'objet de la thèse de Fethi Belghaouti. Une partie de ces travaux a été financée par le projet ANR CAIRE.

**Le cycle de vie de la donnée et la prise de décision** Dans les environnements intelligents, les données ont un cycle de vie précis :

(1) la collecte les données brutes issues des différents capteurs ; (2) l'observation ou le prétraitement de ces données (traitement de l'incomplétude, élimination du bruit et de la redondance, agrégation et normalisation des données) ; (3) le raisonnement pour la reconnaissance des activités et la prise de décision.

Notre contribution s'est faite à différents niveaux : (1) *au niveau perception* en proposant un nouveau modèle, FSCEP (Fuzzy Semantic Complex Event Processing), qui est une extension de CEP (Complex Event Processing). Ce modèle utilise la logique floue et une ontologie de contexte pour enrichir sémantiquement les événements perçus et gérer plusieurs dimensions de l'incertitude, à savoir la fraîcheur, la précision, la consistance et la contradiction des événements. Peu de travaux se sont attaqués à ce défi se limitant souvent à la gestion d'un sous-ensemble des dimensions [37.C.ac] ; (2) *au niveau observation* en introduisant une nouvelle méthode de *segmentation* dynamique qui se distingue de l'état de l'art par sa capacité à adapter la taille de la fenêtre d'observation dynamiquement sans avoir besoin de connaissances préalables. Pour atteindre cet objectif, la méthode combine un processus d'apprentissage et un raisonnement logique basé sur les informations de contexte et les activités de l'utilisateur [3.C.ac]. Les méthodes de *sélection des attributs* classiques ne tiennent pas compte des dépendances et des interactions existantes entre les activités. Pour surmonter cela, une nouvelle méthode de sélection des attributs basée sur l'intégrale de Choquet (FSCI Feature Selection based on Choquet Integral) pour les systèmes de reconnaissance d'activité a été proposée [20.C.ac]. Elle permet de déterminer l'importance d'un attribut et d'analyser les dépendances entre les activités avec une mesure floue. Les résultats expérimentaux démontrent la performance de la méthode proposée par rapport aux méthodes étudiées tant du point de vue taux de reconnaissance que temps de réponse ; (3) *au niveau raisonnement* en proposant des méthodes hybrides combinant de la logique formelle et des techniques d'apprentissage (supervisé et non supervisé) qui, à partir des données de capteurs entachées d'incertitude, infèrent des activités. Des méthodes de propagation de l'incertitude ont été également développées. Les comparaisons avec les solutions de l'état de l'art ont montré de meilleurs taux de reconnaissance malgré la présence des données incertaines [15.C.ac], [14.C.ac]. Cette problématique de reconnaissance d'activité a également été abordée en considérant la distribution des données et du raisonnement. Dans ce cas, des entités réparties observent les données des capteurs, font des prédictions locales, communiquent et collaborent pour identifier les activités en cours. Les

tests expérimentaux montrent une amélioration en termes de précision et de F-mesure par rapport aux approches centralisées [5.C.ac].

Enfin, pour être proactif, un système intelligent doit pouvoir prendre des décisions lorsqu'une situation anormale survient. Pour cela, nous avons développé des algorithmes basés sur l'extraction des causes d'anomalies en utilisant des règles d'association causales et les graphes de liens causaux [1.J.ac] [10.C.ac].

Ces travaux sont menés dans le cadre des thèses d'Amina Jarraya, de Nathan Ramoly et de Hela SFAR (co-encadrée avec Jérôme Boudy de l'équipe ARMEDIA). Certains de ces travaux ont été menés en collaboration avec l'Université de Versailles SQY et l'Université de Tunis et ont été partiellement financés par le projet FUI COCAPS.

**Analyse des réseaux sociaux** Les systèmes de réputation dans les réseaux sociaux ont pour objectif d'aider les utilisateurs à faire la distinction entre les personnes fiables des malveillantes en leur associant des valeurs de confiance. Cette connaissance des liens entre les participants des réseaux sociaux peut être utile pour les systèmes de recommandation. Nous avons proposé un nouvel algorithme de mesure de réputation basé sur un réseau de confiance. Cet algorithme classe les utilisateurs selon leurs similarités de confiance directe (entre deux nœuds voisins) et indirecte (entre nœuds distants) que nous mesurons à l'aide de deux modèles que nous avons développés. Des évaluations ont été menées sur des données réelles de Twitter, avec plus de 250 000 utilisateurs et 320 000 relations. Les résultats démontrent que nos modèles génèrent de meilleurs résultats, en termes d'utilité et d'efficacité, que les approches pionnières existantes. Par ailleurs, nous avons enrichi le vocabulaire de l'ontologie FOAF (Friend Of A Friend) avec la réputation des utilisateurs pour satisfaire leurs besoins en stockant des données de confiance importantes dans une structure bien organisée et facile à comprendre. Ces travaux ont été initiés dans le cadre de la thèse de Sana Hamdi [12.J.ac] et se poursuivent avec la thèse de Monika Rakoczy [27.C.ac]. Certains de ces travaux ont été réalisés en collaboration avec l'Université de Tunis.

**La collecte participative** Le crowdsensing mobile (MCS) est devenu un paradigme prometteur pour faciliter les applications de monitoring urbain telles que la surveillance de l'environnement, la détection de congestion au niveau du trafic ou le partage d'informations publiques. Le MCS exploite la mobilité des utilisateurs, des capteurs embarqués dans les téléphones mobiles et l'infrastructure sans fil existante pour capturer et collecter les données environnementales, permettant de détecter à moindre coût diverses données urbaines dans des régions non couvertes d'infrastructures spécialisées. Le défi est alors de réduire le coût de détection en ne considérant qu'un petit nombre de sous-zones tout en garantissant un niveau satisfaisant de qualité des données pour l'ensemble de la zone cible. Notre approche consiste à détecter une partie des sous-zones par les participants, tandis que les données du reste des sous-zones sont inférées [23.C.ac]. La faisabilité théorique réside dans le fait qu'il existe des corrélations spatio-temporelles élevées dans la plupart des données urbaines (ex. la qualité de l'air et le bruit). [49.C.ac, 9.J.ac, 29.C.ac, 6.C.ac]. En plus des données collectées par les participants, d'autres sources sont aussi fusionnées pour augmenter les performances de l'inférence (ex. les points d'intérêt de Google Places, la météo et le trafic routier sont ajoutés pour la mesure de la pollution atmosphérique). Nous modélisons l'inférence de données comme un problème de complétion de matrice. Pour une allocation optimale des tâches, nous utilisons plusieurs algorithmes d'inférence, comme les  $k$  plus proches voisins (k-NN) et la technique de l'acquisition comprimée (compressive sensing) [38.C.ac, 18.C.ac]. Pour l'évaluation de la qualité des données, par la validation croisée "leave-one-out" sur les cellules détectées, nous avons obtenu plusieurs échantillons de l'erreur d'inférence. Sur la base de ces échantillons, nous utilisons l'inférence bayésienne pour estimer la fonction de densité de probabilité de la précision d'inférence. D'autres solutions de crowdsensing ont été proposées pour optimiser différents critères, comme le nombre de capteurs ou l'énergie afin d'atteindre une couverture et une qualité de données [22.J.ac, 5.J.ac, 10.J.ac, 34.C.ac, 33.C.ac].

Certains de ces travaux ont été menés en collaboration avec l'Université de Pékin, la *School of Computer Science, Northwestern Polytechnical* de Chine et l'*University of Science and Technology* de Hong Kong.

### 3.2.2 Faits marquants de l'équipe ACMES

- **Structuration de l'équipe** : Intégration de l'axe Parallélisme et Système
- **Récompenses** : "Best Paper Award" à UIC 2015, IEEE CPSCOM 2013, MobiQuitous 2017 et PICOM 2017 ; "Best Student Paper Award" à SCC 2018 ; "Distinguished Paper Award" à CAiSE 2018 ; "Top 5 Best Paper Award" à SCC 2016 ; "Honorable Mention Award" dans UbiComp 2015 et UbiComp 2016 ; Dingqi Yang a obtenu le prix de la meilleure thèse Samovar ainsi que le prix de thèse chinois Excellent Overseas Chinese



PhD Thesis Award en 2015.

— **Organisation de conférences :**

- ICSOC 2014 conférence phare du domaine des service
- Workshop "Middleware for IoT (M4IoT)" à la conférence ACM Middleware (de 2014 à 2018)

— **Réalisations logicielles et plateformes :**

- La plateforme logicielle M4IOT fournit un ensemble de logiciels pour l'Internet des objets<sup>7</sup>. Cette plateforme est utilisée notamment pour développer des démonstrateurs dans le cadre de la plateforme Hadaptic<sup>8</sup> pour l'assistance aux personnes dépendantes et pour la ville intelligente, en particulier dans le cadre du projet ANR INCOME<sup>9</sup>.
- Depuis 2011, l'équipe ACMES développe et maintient le logiciel open-source d'analyse de performance pour les applications parallèles EZTrace. Ce logiciel est utilisé par les communautés académique et industrielle et a été téléchargé plusieurs centaines de fois. Les projets FUI IDIOM (démarrant en septembre 2018) et ANR JCJC Pythia (2018) reposent en grande partie sur EZTrace.
- Plus récemment, l'équipe ACMES a développé le logiciel open-source d'analyse mémoire NumaMMA dans le cadre d'un projet DIM-RFSI.
- L'équipe ACMES développe depuis 2017 l'hyperviseur pour machines NUMA Scalevisor. Cet hyperviseur est la brique de base du projet ANR PRCE Scalevisor (2018) qui vise à permettre l'exécution efficace d'applications dans le cloud grâce au rack-scale computing.
- Dans le cadre du projet H2020 CloudButton, l'équipe ACMES en collaboration avec RedHat participe au développement du système de stockage Infinispan<sup>10</sup>.
- L'équipe ACMES participe au développement d'Apache Gora<sup>11</sup>.

### 3.3 Organisation et vie de l'équipe ACMES

La vie de l'équipe s'organise autour de réunions régulières et de séminaires de chercheurs invités. Les doctorants de l'équipe sont incités à présenter leurs travaux à ces occasions. Depuis 2013, l'équipe ACMES a invité plus de 20 chercheurs extérieurs pour des séminaires parmi lesquels :

- From Scientific Process Management to Process Science par Jan Mendling (WU Vienna), nov. 2015.
- The Development of IoT Applications : Requirements and state of the art par Thais Batista (UFRN, Brésil), déc. 2015.
- A Short History of SLAs : Why service quality matters more now and why it is more difficult par Heiko Ludwig (IBM Almaden Research Center), juin 2016.
- Debugging applications : a data mining approach par Vincent Leroy (Université de Grenoble), mai 2017.
- Research Agenda on Enterprise 2.0 @ Zayed University : Past, Present, and Future par Zakaria Maamar (Zayed University in Dubai), juillet 2017.
- Software Heritage : la bibliothèque d'Alexandrie du code source par Roberto di Cosmo (INRIA), nov. 2017.
- SafeFS : A Modular Architecture for Secure User-Spac File Systems (One FUSE to rule them all) par Valerio Schiavoni (Université de Neuchatel), janv. 2018.
- Case Management in the Age of Data-driven Insights and Cognitive Services par Boualem Benatallah (UNSW Sydney), avril 2018.
- Vers de nouveaux défis en apprentissage statistique et en science des données par Mustapha Lebbah (Université Paris 13), mai 2018.
- From the Stability of formal concepts to minimal transversals : fundamentals and applicative cases par Sadok Ben Yahia (Faculté des Sciences de Tunis), mai 2018.

---

7. [www-inf.telecom-sudparis.eu/m4iot](http://www-inf.telecom-sudparis.eu/m4iot)

8. [hadaptic.telecom-sudparis.eu](http://hadaptic.telecom-sudparis.eu)

9. <http://www-inf.it-sudparis.eu/MARGE/Showcase/VideoINCOME-vFrancaise.mp4>

10. ([Infinispan.org](http://Infinispan.org))

11. ([gora.apache.org](http://gora.apache.org))

### 3.4 Analyse SWOT de l'équipe ACMES

Depuis notre précédente évaluation de 2013, l'équipe a vu son effectif s'accroître, avec l'arrivée de cinq enseignants-chercheurs sur la thématique parallélisme et système. Cette nouvelle thématique procure à l'équipe une dimension système essentielle au développement d'applications en fournissant la puissance nécessaire aux calculs via l'utilisation efficace des technologies matérielles les plus performantes aujourd'hui (multicœur, accélérateurs matériels, racks etc.).

A la lumière de ces évolutions et des résultats, on peut effectuer l'auto-analyse suivante :

Points forts	<ul style="list-style-type: none"> <li>— Bonne production scientifique. Plusieurs best paper awards dans des conférences du domaine ;</li> <li>— Reconnaissance nationale et internationale : organisation de conférences, dissémination de logiciels, participation à des comités de programme et à des comités de rédaction de revues, publications dans les conférences et revues de référence de nos domaines ;</li> <li>— Constitution d'une équipe de taille importante intégrant des compétences complémentaires (systèmes répartis, services et cloud, gestion de données et de connaissances, IA, performance) ;</li> <li>— ACMES participe activement à la mention Calcul Haute Performance et Simulation et aux Master2 CILS et Datascale de la mention informatique de l'Université Paris-Saclay sur des thématiques bien en phase avec celles de l'équipe, ainsi qu'à la formation de chercheurs et doctorants au sein de l'école doctorale STIC de l'Université Paris-Saclay ;</li> <li>— Réseau de collaborations à l'international (Allemagne, Brésil, Espagne, Chine, Suisse, Tunisie) nous amenant des stagiaires, doctorants et post-doctorants de qualité.</li> </ul>
Points faibles	<ul style="list-style-type: none"> <li>— Manque de ressources permanentes en ingénieur plateforme, dommageable pour nos thématiques de recherche ;</li> <li>— Effort de collaborations entre les thématiques de l'équipe à poursuivre et amplifier ;</li> </ul>
Opportunités	<ul style="list-style-type: none"> <li>— Le futur déménagement partiel du laboratoire vers le plateau de Saclay (2019) aidera l'équipe à se rapprocher d'acteurs académiques et industriels avec qui nous avons déjà des contacts (LRI, LTCI, ENSTA, EDF R&amp;D, etc.). Ceci facilitera encore plus les collaborations et permettra à l'équipe de travailler dans un cadre scientifique encore plus favorable.</li> <li>— Des domaines porteurs : Intelligence artificielle (Institut de convergence DATAIA, futur appel 3IA), IoT, Cloud-infrastructures et Industrie du Futur (programme franco-allemand)</li> </ul>
Risques	<ul style="list-style-type: none"> <li>— Dispersion de l'équipe due à un grand nombre de thématiques</li> <li>— La bi-localisation Evry-Plateau de Saclay difficile à organiser</li> </ul>

### 3.5 Projet scientifique de l'équipe ACMES

Le projet du précédent quinquennat était centré sur l'aspect multi-échelle dans l'ubiquitaire, sur l'aspect formalisation et données dans le cloud et sur l'aspect social et hybridation de modèles pour l'intelligence collective et sociale. Ces thèmes restent d'actualité mais vont évoluer pour prendre en compte les nouveaux usages de l'informatique.

Dans les prochaines années, l'informatique va être de plus en plus utilisé pour nous aider et nous accompagner dans notre vie quotidienne. En particulier, les domaines de la simulation (e.g., modélisation de médicaments ou de phénomènes physiques), de la prédiction (e.g., prédire la consommation électrique d'un quartier avec les compteurs linky ou la meilleure façon d'agencer les produits sur la page web d'une boutique en ligne pour maximiser les profits) ou de l'automatisation (e.g., l'industrie du futur) sont en plein essor. Dans ces domaines, les infrastructures logicielles et matérielles sont caractérisées par la présence de nombreux nœuds de calculs, que ce soit de petits

appareils pour collecter les données ou de gros serveurs pour les traiter. Ces infrastructures sont aussi caractérisées par une forte complexité des architectures matérielles et des piles logicielles. Ces infrastructures sont soumises à une haute dynamique, qu'elle soit due à l'arrivée ou au départ intempestif de nœuds de calculs, à des changements de comportements des utilisateurs entraînant des charges de travail variables, ou à des mises à jour à la volée du logiciel. Ces infrastructures doivent enfin être capable d'analyser et d'extraire des connaissances en quasi temps réel à partir de volumes colossaux de données issues de sources hétérogènes et produites à des vitesses rapides et variables afin de diagnostiquer rapidement et précisément les phénomènes normaux ou anormaux sous-jacents.

Dans le prochain quinquennat, l'équipe ACMES compte s'intéresser à trois challenges associés aux caractéristiques de ces infrastructures matérielles et logicielles : (1) le *passage à l'échelle*, (2) la *complexité* des architectures matérielles et des piles logicielles et (3) le caractère *dynamique* des infrastructures et des données massives et hétérogènes véhiculées par ces infrastructures.

Vu l'expertise de ses membres, l'équipe ACMES s'intéressera à ces challenges en suivant trois grands axes :

**Performance** : la simulation et la prédiction nécessitent une importante puissance de calcul et il est important d'améliorer les performances des infrastructures logicielles pour éviter de gaspiller les ressources matérielles et l'énergie.

**Génie logiciel et intergiciel** : la complexité, l'hétérogénéité et la dynamique inhérente aux infrastructures matérielles et logicielles rendent les codes difficiles à écrire et à maintenir. Il devient donc nécessaire de proposer de nouveaux intergiciels et modèles pour simplifier la conception des applications.

**Données et connaissances** : automatiser ou effectuer des prédictions efficaces nécessite de collecter et de traiter de grands volumes de données hétérogènes, et il est donc nécessaire de proposer et d'étudier de nouveaux algorithmes permettant de collecter et traiter efficacement ces données.

L'équipe ACMES traitera ces trois axes à tous les niveaux de la pile logicielle : au niveau système, au niveau intergiciel et au niveau applicatif. La suite de la section présente plus précisément les sujets que nous comptons aborder à chaque niveau de la pile logicielle.

### 3.5.1 Parallélisme et système

Dans le monde des grands serveurs, que ce soit celui du cloud computing ou celui du calcul haute performance, les architectures matérielles sont en profonde mutation. Les processeurs récents sont devenus de véritables systèmes distribués avec des hiérarchies complexes de caches et des cœurs connectés par des réseaux complexes (e.g., la topologie mesh du INTEL Skylake<sup>12</sup>). Ces processeurs offrent régulièrement de nouveaux jeux d'instructions, comme le jeu d'instructions SGX introduit par INTEL Haswell permettant de crypter/décrypter à la volée les données ou le jeu d'instructions INTEL DCAT permettant de configurer finement les hiérarchies de caches. Une machine de classe serveur peut facilement agréger entre 4 et 8 de ces processeurs et les interconnecter avec les périphériques via un second réseau tout aussi complexe (architectures NUMA). Les hiérarchies de mémoires et de stockages sont aussi en profonde mutation avec l'arrivée des disques NVMe à débit ultra-rapide et des mémoires non volatiles (NVRAM) comme la DIMM Optane récemment annoncée par INTEL<sup>13</sup>. Cette évolution des architectures matérielles commence aussi à s'étendre à l'échelle de plusieurs machines avec le mouvement autour du rack-scale computing et de la désagrégation<sup>14 15</sup> qui, grâce à des réseaux ultra-rapides comme Infiniband, devrait permettre à terme à n'importe quelle machine d'un rack d'utiliser n'importe quelle ressource (mémoire, périphérique) d'une autre machine.

Durant les prochaines années, le groupe parallélisme et système va s'intéresser à l'impact de ces évolutions sur le logiciel en suivant trois axes.

Le premier axe concerne l'étude des abstractions que doit fournir un système pour rendre l'utilisation de ces architectures simples et efficaces. Dans ce cadre, nous allons nous intéresser à deux thèmes : la sécurité avec SGX en proposant de nouvelles abstractions simplifiant la programmation d'applications respectant la vie privée (projet ANR PCRI Primate démarrant en juin 2018), et aux mémoires non volatiles en proposant de nouvelles structures de données permettant de rendre les *object storages* plus efficaces et plus simples à programmer (thèse de A. Lefort démarrant en septembre 2018 et thèse de R. Dulong en co-tutelle avec l'université de Neuchatel démarrant en septembre 2018).

12. <https://itpeernetwork.intel.com/intel-mesh-architecture-data-center/>

13. <https://www.anandtech.com/show/12828/intel-launches-optane-dimms-up-to-512gb-apache-pass-is-here>

14. <https://www.microsoft.com/en-us/research/project/rack-scale-computing/>

15. <https://www.intel.com/content/www/us/en/architecture-and-technology/rack-scale-design-overview.html>

Le second axe concerne l'amélioration des performances et de la consommation énergétique des piles logicielles patrimoniales. Dans ce cadre, nous allons nous intéresser à l'utilisation de techniques de profilage pour exploiter plus efficacement ces architectures matérielles complexes (thèse de A. Daumen en co-tutelle avec le CEA démarrant en septembre 2018, projet ANR JCJC 2018 Pythia, projet FUI Idiom démarrant fin 2018), à l'analyse, la modélisation et l'optimisation de la consommation d'énergie d'applications de calcul sur ces architectures complexes (travaux de A. Guermouche en collaboration avec A. C. Orgerie à l'IRISA), à améliorer les performances des runtimes permettant l'analyse de grands volumes de données sur ces architectures complexes (travaux de E. Brunet en collaboration avec le Genopole), et à l'utilisation de la virtualisation pour cacher la complexité du matériel au logiciel dans le cadre du rack-scale computing (thèse de A. Lescouet démarrée en septembre 2017 et projet ANR PRCE 2018 Scalevisor).

Le troisième axe concerne l'impact de la désagrégation et de la géo-distribution des données sur les architectures orientées stockage. Son étude a commencé par des travaux autour des systèmes de fichiers à cohérence ajustable en coopération avec Scality, l'Université de Grenoble et l'INRIA Paris (thèse de Tuanir F. Rezende démarrée en août 2017 et projet ANR RainbowFS financé par l'instrument PRCE 2017). Elle se poursuivra autour de travaux portant sur l'architecture de type Function-as-a-Service (FaaS), qui simplifie le chargement et l'exécution massivement parallèle de code dans le Cloud (projet H2020 accepté à l'appel CT-12:2018-2020).

### 3.5.2 Génie logiciel et intergiciel

Depuis de nombreuses années, le génie logiciel et les intergiciels, mettent à la disposition des développeurs les méthodes et les cadres logiciels pour faciliter la gestion du cycle de vie (conception, développement, test, déploiement, exécution) des systèmes de plus en plus complexes tout en assurant à ces systèmes des qualités de portabilité et d'interopérabilité. Les systèmes de demain seront toujours de plus en plus complexes notamment à cause de la dispersion géographique des briques le composant, et du besoin de passage à l'échelle. Ce qui pose de nombreux défis en termes d'intergiciels et de génie logiciel. Nous projetons de focaliser nos efforts sur deux points de vue particuliers : l'Internet des objets ainsi que l'orchestration et l'élasticité des ressources Cloud.

#### Génie logiciel et intergiciel pour l'Internet des Objets

Les prochaines années vont voir se développer de nouveaux services tirant profit de l'Internet des objets. L'Internet des objets pose de nouveaux défis en termes de génie logiciel pour le développement de nouveaux services. Le découplage des objets connectés et des applications qui y accèdent, l'hétérogénéité des systèmes, la dynamique et la mobilité des objets connectés, le volume de données produites en continu sont autant de caractéristiques des systèmes dans le contexte de l'IoT. Nous prévoyons de poursuivre nos travaux entamés sur les intergiciels et le génie logiciel au service des concepteurs d'applications dans le contexte de l'Internet des objets. Une de nos préoccupations est de procurer le bon niveau d'abstraction aux concepteurs d'applications pour l'IoT, tout en concevant les briques intergicielles qui répondent aux défis de l'Internet des objets. Une autre préoccupation est de proposer des intergiciels qui assurent des propriétés de passage à l'échelle, de sûreté de fonctionnement et de sécurité. Nous citons de manière non exhaustive quelques axes sur lesquels nous projetons de travailler dans les prochaines années : les services de découverte à l'échelle de l'IoT, la distribution des informations produites par l'IoT, et la protection de la vie privée.

Le premier axe concerne le service de découverte pour l'IoT. Un des défis majeurs de l'IoT vient à la fois du nombre d'objets connectés et du volume de données à filtrer et traiter dans un temps contraint. Dans ce contexte, nous proposons de concevoir des services de découverte à très grande échelle pour la sélection d'objets et de données. Ces services permettront la sélection des données ayant un niveau de qualité satisfaisant tout en respectant la vie privée des utilisateurs. D'un point de vue génie logiciel, nous envisageons le couplage de ces services de sélection avec des intergiciels côté client qui s'interfacent de manière transparente avec la plateforme IoT ou le service de découverte. L'objectif est d'offrir le bon niveau d'abstraction côté applicatif tout en assurant la gestion des interactions avec les services de découverte de manière transparente et économique du point de vue de l'utilisation de la batterie. Des travaux ont démarré dans cet axe avec l'Université Fédérale du Rio Grande do Norte (UFRN), Natal, Brésil

Le deuxième axe concerne la distribution d'informations de contexte en provenance de l'IoT à différentes échelles (nuage, géorépartition, etc.). Les systèmes répartis construits avec le paradigme de l'Internet des objets sont fortement découplés : découplage en termes de temps, d'espace et de synchronisation, des consommateurs, des

producteurs et des unités de traitement. Ces caractéristiques accroissent les problèmes de nommage, de découverte, de filtrage, de routage, de déploiement à la demande, de tolérance aux fautes et aux déconnexions, et de passage à l'échelle. Nos travaux ont commencé à aborder ce défi par la différenciation de la dissémination des informations, du placement des traitements selon les dimensions et les échelles considérées. Les premières dimensions et échelles prises en compte concernent les objets connectés avec faible capacité de communication et de traitement ; passerelles et serveurs de proximité géorépartis ; serveurs de calcul en nuage ; dispositifs des utilisateurs. Nous prévoyons de continuer l'étude des systèmes multi-échelles en ajoutant la prise en compte de la forte dynamique due à la nomadicité des clients ainsi qu'à la mobilité des objets et des unités de traitement.

Le troisième axe concerne la protection de la vie privée. La protection des données personnelles est aujourd'hui l'un des défis majeurs de la mise en œuvre de l'Internet des objets pour créer la confiance nécessaire à l'acceptation de nouveaux usages par les utilisateurs. Des travaux concernant plus particulièrement la gestion du consentement de manière sécurisée ainsi que le contrôle d'usage ont commencé en collaboration avec l'équipe ACES du laboratoire LTCl, et avec l'équipe R3S du laboratoire SAMOVAR.

### **Orchestration et élasticité des ressources Cloud**

La clé de voute pour exploiter pleinement le potentiel du Cloud Computing est de maîtriser l'orchestration et l'élasticité des ressources dans le développement des applications. L'orchestration désigne le processus de sélection, de déploiement, de surveillance, et de contrôle d'exécution des ressources avec une QoS souhaitée. L'élasticité est définie comme le pouvoir de reconfigurer dynamiquement la capacité des ressources Cloud. Les approches et les outils existants (Docker, Terraform, Puppet, Juju, Ansible, Amazon OpsWorks, Chef etc.) pour supporter une orchestration automatisée, tout en maîtrisant directement les fonctionnalités d'élasticité, sont encore insuffisants car (i) ils s'appuient généralement sur différents langages de programmation procéduraux et des langages de script de bas niveau, (ii) chaque outil repose sur des modèles de description de ressources, des interfaces de gestion et d'accès différents. Par conséquent, même des programmeurs experts sont obligés de maîtriser différentes APIs de bas niveau, une syntaxe de ligne de commande et différents langages de programmation pour décrire les configurations de ressources Cloud et supporter les caractéristiques d'élasticité. Nous pensons que les caractéristiques d'élasticité devraient être fournies au niveau de la description des ressources. Au lieu de s'appuyer sur des mécanismes de script de bas niveau ou les règles spécifiques à l'outil utilisé, nous soutenons que les modèles et les langages de description des ressources Cloud devraient être dotés de constructions intuitives permettant de spécifier une gamme de mécanismes d'élasticité flexibles. Notre objectif est donc de proposer un nouveau cadre d'orchestration des ressources Cloud qui intègre la conception et la vérification des fonctions d'élasticité et qui peut s'étendre à divers outils d'automatisation en adoptant une méthodologie dirigée par les modèles. Plus précisément, le cadre proposé vise à fournir quatre fonctionnalités principales :

1. Modélisation des ressources Cloud : cela consiste à modéliser à la fois les ressources et leurs caractéristiques d'élasticité. Pour modéliser les ressources Cloud, leurs relations et tous les artefacts d'implémentation requis, nous adoptons TOSCA, la norme pour la spécification de topologie et d'orchestration pour les applications Cloud. En outre, nous allons adopter un formalisme basé sur une machine à états pour capturer les caractéristiques d'élasticité en tirant profit de l'intuitivité et de l'efficacité de la représentation de l'aspect comportemental,
2. Vérification de l'élasticité : cette fonctionnalité vise à vérifier le comportement de la machine à états définie avant de la faire correspondre à des artefacts spécifiques nécessaires pour le contrôle d'exécution. Pour ce faire, la machine à états est transformée en automates temporisés sur lesquels des propriétés comportementales sont vérifiées,
3. Génération d'artefacts spécifiques : en adoptant une approche dirigée par les modèles, on traduit automatiquement les artefacts génériques des ressources Cloud (conçus avec TOSCA) vers des artefacts spécifiques à des outils,
4. Orchestration de ressources Cloud : l'orchestration est basée sur un ensemble de connecteurs qui assurent l'interaction avec les divers outils. Plus particulièrement, en exploitant à la fois les artefacts générés à l'étape 3 et ceux fournis publiquement par les communautés des outils, ces connecteurs offrent des opérations d'orchestration de bout en bout, notamment la sélection, le déploiement, la surveillance et le contrôle d'exécution.

## Procédés métiers et Blockchain pour l'Internet des Objets

Bien que la communauté des processus métiers considère  $\acute{n}$  l'activité  $\acute{z}$  comme l'unité de base d'un processus métier, la prochaine tendance serait de permettre à des objets d'être au cur d'un processus métier en représentant des activités auto exécutables et qui peuvent former des réseaux de collaboration ad hoc et multi tenants. Dans ce contexte, notre objectif est de mettre en lumière un nouveau paradigme appelé  $\acute{n}$  processus d'objets  $\acute{z}$ , qui est une nouvelle manière de relier l'Internet des Objets et l'univers des processus métiers. Ce paradigme s'assurera que les objets ne fonctionneront pas comme des silos mais contribueront collectivement à offrir des services à valeur ajoutée aux entreprises telles que développer des applications agiles autour des objets connectés et accéder à plus de clients par des objets adaptables.

Les relations prédéfinies entre les objets (par exemple, proximité et voisinage) soutiendront le développement de réseaux des contacts (privilegiés) de sorte qu'un objet puisse : (i) contacter d'autres objets via ces réseaux, (ii) éviter les conflits avec d'autres objets faisant déjà partie du processus, et (iii) recommander d'autres objets pour les inclure ou exclure des processus en fonction des expériences passées. Dans le paradigme de processus d'objets, les capacités (ex., détecter, stocker, traiter et diffuser) prescriront les fonctions d'un objet une fois qu'il devient actif et les préparent, ainsi, à constituer des groupes de collaboration avec d'autres objets. Un bon nombre de défis et questions ouvertes en découlent incluant la définition d'une architecture d'objets sociaux, l'interopérabilité sémantique des objets, la découverte des objets, l'évaluation de la confiance des objets, pour n'en citer que quelques-uns.

Par ailleurs, les défis de la conception conjointe et le manque de confiance mutuelle peuvent entraver une plus large adoption des  $\acute{n}$  processus d'objets  $\acute{z}$  qui sont par nature multi tenants. La technologie émergente de Blockchain a le potentiel de changer radicalement l'environnement dans lequel les processus multi tenants peuvent fonctionner. Ainsi, la technologie Blockchain permet de repenser l'exécution des processus métier inter-organisationnels en raison de leur potentiel à la réaliser sans un orchestrateur central qui sert d'un point de confiance (et de défaillance). Concrètement, une grande partie du flux de contrôle et la logique métier des processus métier multi tenants peut être compilés à partir de modèles de processus dans des contrats intelligents (smart contracts) qui garantissent que le  $\acute{n}$  processus d'objets  $\acute{z}$  est correctement exécuté. Les composants dits déclencheurs permettent de connecter ces implémentations de processus inter-organisationnelles à des services Web, des objets, et à des implémentations de processus internes. Ces déclencheurs servent de pont entre la blockchain et les applications d'entreprise.

Ces fonctionnalités démontrent comment Blockchain peut aider les organisations à mettre en uvre et exécuter des processus même s'ils ne peuvent pas se mettre d'accord sur un tiers de confiance. Cependant, des réponses doivent être proposées aux problèmes liés à l'expressivité et la richesse sémantique des contrats intelligents, ainsi que l'évolution de ce modèle de chorégraphie pour refléter les modifications apportées par une nouvelle version du modèle de processus.

### 3.5.3 Données et Connaissances

L'axe Intelligence Collective et Sociale a évolué progressivement vers une thématique plus générique sur les *Données et Connaissances* qui renvoie aux thématiques de recherche sur la gestion des données, le cycle de vie des données, la sémantique des données, l'extraction de connaissances et la modélisation des raisonnements afin de manipuler ces connaissances. Cet axe s'inscrit dans un domaine plus générique des *Sciences de Données* et de l'*Intelligence Artificielle*.

De plus en plus d'applications (ville intelligente, trafic urbain, monitoring, réseaux sociaux) requièrent un traitement en temps réel des flux de données pour répondre à des requêtes souvent complexes nécessitant la fusion de plusieurs sources hétérogènes et généralement non fiables. Un système capable de répondre à ces requêtes doit être en mesure de satisfaire ces exigences : manipuler un volume de données important, traiter les flux de données à la volée, gommer l'hétérogénéité et prendre en compte l'incertitude inhérente aux données. Des systèmes capables de traiter en temps réel des flux de données volumineux existent (gestion d'événements complexes, gestion de flux de données) mais ces derniers ne couvrent pas l'ensemble des exigences que nous nous fixons.

Pour satisfaire ces exigences, nous envisageons de poursuivre nos travaux sur les flux de données sémantisés et d'aborder différents aspects de la problématique selon quatre volets complémentaires : le premier va s'intéresser au traitement du flux et en particulier le problème de la relaxation de requêtes sur les flux de données RDF, le second va plutôt s'intéresser au raisonnement continu sur les flux de données RDF pour améliorer les réponses aux requêtes et faire des prédictions, le troisième va s'intéresser à l'extraction et l'analyse en temps réel de réseaux

sociaux en flux et enfin, le quatrième va aborder un problème qui se situe en amont des précédents, à savoir la fusion de données hétérogènes.

**La relaxation de requêtes sur les flux de données RDF** Dans ce premier volet, nous nous intéressons au traitement du flux et plus spécifiquement à la problématique de relaxation de requêtes. C'est une approche qui consiste à transformer une requête en situation d'échec en une nouvelle requête dont les conditions exprimées sont relâchées pour augmenter les chances de retourner des réponses à l'utilisateur. Ce problème n'est pas nouveau mais nécessite d'être revisité pour mieux l'adapter aux propriétés des données au format RDF et produites en flux. Le premier objectif que nous visons est de proposer un encodage du problème de relaxation de requêtes sur des flux sémantiques comme un problème de satisfiabilité propositionnelle, par l'exploitation de concepts issus de l'intelligence artificielle symbolique. Le deuxième objectif est de fournir des explications satisfaisantes pour les causes d'échec de la requête exprimée par l'utilisateur. Il s'agit ici de définir des notions d'explications acceptables et rationnelles. Ces travaux ont été initiés avec la thèse de W. Mebrek en cotutelle démarrée en octobre 2017.

**Le raisonnement continu sur les flux de données RDF** Ce deuxième volet concerne la partie raisonnement sur les flux. Ce domaine, issu de la combinaison des techniques de raisonnement et de la gestion de flux de données, permet d'améliorer les résultats de calcul pour la réponse aux requêtes et la prise de décision sur les données préalablement agrégées ou fusionnées et filtrées. Nous projetons de développer des modèles logiques expressifs de représentation des flux de données incertaines. Ces modèles doivent être capables de prendre des décisions raisonnables et intelligentes en fonction d'objectifs assignés. La plupart des approches existantes sont basées sur des langages logiques (souvent des extensions de OWL2 DL) mais ces derniers sont peu adaptés à la manipulation de données ayant des fréquences élevées. Le défi ici réside dans le fait de trouver un compromis entre la complexité de la méthode de raisonnement et la fréquence du flux de données que le raisonneur est capable de gérer sans rencontrer de problèmes. Nous souhaitons utiliser d'autres formalismes plus expressifs pour le raisonnement déductif comme les logiques temporelles et plus généralement des alternatives à la logique de description. De plus, il serait intéressant d'explorer l'intégration du raisonnement déductif avec d'autres types de raisonnements comme le raisonnement inductif. D'une façon plus générale, les questions suivantes se posent : Comment plusieurs techniques de raisonnement peuvent-elles coexister ? Est-il possible d'exploiter des paradigmes plus interactifs dans lesquels les résultats de méthodes d'apprentissage et les techniques de raisonnement coopèrent pour atteindre un objectif donné ? La thèse CIFRE de M. Bilani avec Ozwillo (<https://www.ozwillo.com/en>) qui démarre fin 2018 a pour ambition de répondre à ces interrogations.

**Analyse des réseaux sociaux en flux** Nous prévoyons de poursuivre nos travaux dans ce troisième volet sur l'analyse des réseaux sociaux en ajoutant les mêmes contraintes de dynamique que les deux premiers volets. Les réseaux sociaux en flux sont des réseaux où des mises à jour rapides et continues surviennent aussi bien au niveau de la structure elle-même qu'au niveau des données associées aux nœuds et aux liens. Reasonner sur ces données en continu et en tirer des informations significatives et non triviales ou implicites en temps réel est également un défi majeur dans de nombreuses applications comme l'analyse de sentiments, la détection de comportements malveillants, les recommandations sociales, les tendances personnalisées, la détection de spam, etc. La nature complexe des réseaux sociaux et leur évolution rapide soulèvent de nombreux problèmes qui n'ont pas été abordés dans la littérature. Par exemple, comment permettre à un utilisateur d'effectuer une grande variété de requêtes d'analyse sur des flux de données en temps réel. Ces requêtes peuvent aller de simples requêtes d'abonnement/souscription pour recevoir des notifications, à des requêtes de détection d'anomalies complexes, telles que l'identification en temps réel d'un comportement anormal. Les solutions apportées dans les deux précédents volets constitueront une base pour répondre aux nouveaux défis soulevés par les réseaux sociaux en flux. Des travaux ont déjà été initiés dans [13.C.ac] pour la détection de communauté en collaboration avec le CRIL, Université d'Artois.

**La fusion de données hétérogènes** Enfin, une problématique transverse aux trois volets précédents est la fusion de données hétérogènes. C'est une étape critique dans l'intégration des données collectées via plusieurs capteurs. Les données provenant de différentes sources et appartenant à différents domaines (e.g. données de trafic, media sociaux, qualité de l'air), ont une représentation, une distribution, une échelle et une fréquence différentes. De plus, elles ne sont pas fiables. La fusion de données hétérogènes vise à tirer parti de la puissance des données

multimodales ou spatio-temporelles et à proposer de nouveaux modèles et algorithmes pour améliorer l'inférence du contexte et la précision des prévisions. C'est un problème difficile en l'absence de liens sémantiques explicites entre ces données. Par exemple, si l'on veut calculer la moyenne des températures produites par plusieurs capteurs, il serait absurde d'aggréger les données provenant d'une centrale nucléaire et celles relevées dans une habitation. Une solution potentielle à ce problème serait de ramener la fusion de donnée à une fusion de connaissances s'il on dispose d'ontologies décrivant les contextes de production de ces données.



# 4 Équipe ARMEDIA

## 4.1 Présentation de l'équipe ARMEDIA

### 4.1.1 Introduction

L'équipe ARMEDIA a été constituée en janvier 2016 suite à la restructuration des forces de recherche de Telecom SudParis autour de l'Université Paris Saclay et de ses écoles doctorales. Outre des chercheurs de SAMOVAR précédemment affiliés à l'équipe TIPIC, cette nouvelle équipe inclut une composante (4 chercheurs) rattachée initialement au laboratoire MAP5 de Paris Descartes (UMR 8145) en tant qu'équipe *Ingénierie Multimédia* et qui ne pouvait pas être maintenue en l'état par rapport à la politique de regroupement des Universités et d'affiliation aux écoles doctorales. Le CNRS a donné son accord pour une bi-affiliation de cette composante jusqu'à l'évaluation HCERES de MAP5 (20 février 2018, Vague D).

La nouvelle équipe ARMEDIA de SAMOVAR réunit des chercheurs partageant des intérêts scientifiques proches, notamment sur une recherche appliquée autour des données multimédia, en incluant la santé et l'assistance à la personne dépendante, d'où l'acronyme de l'équipe, *Applied Research for (multi)Media Enrichment, Diffusion, Interaction and Analysis*.

L'équipe est implantée sur deux sites, à Telecom SudParis, Evry et au CEA NanoInnov, Saclay, ce qui est précurseur de la future bilocalisation de Telecom SudParis sur Evry et Saclay et nous a imposé d'adapter les actions d'animation.

### 4.1.2 Effectifs et moyens de l'équipe

Dénombrant 9 chercheurs au démarrage, ARMEDIA s'est renforcée courant 2016 avec l'affectation d'une Maître de Conférences de Telecom SudParis et d'un chercheur émérite CNRS (anciennement membre du laboratoire LTCI de Telecom ParisTech). L'équipe compte également deux chercheurs associés (Telecom SudParis et ESME Sudria) dont le statut pourrait évoluer à terme vers membre permanent. A noter que seulement l'activité des membres permanents de l'équipe est présentée pour cette évaluation, comme suit : pour les membres SAMOVAR ayant rejoint l'équipe ARMEDIA, la période concernée est au plus 2013-2018 (en fonction de la date d'entrée à SAMOVAR) ; pour les anciens membres MAP5 on va distinguer deux périodes, avant/après affiliation à SAMOVAR (2013-2015 MAP5 et 2016-2018 SAMOVAR). Les données financières seront répertoriées seulement depuis la création d'ARMEDIA (2016, 2017).

Le tableau suivant synthétise les effectifs de l'équipe sur la période 2013-2016, les membres permanents étant comptés à partir de 2016 uniquement, date de création de l'équipe.

	2013	2014	2015	2016	2017	→ 30/6/2018
Enseignants chercheurs permanents				11	11	11
Chercheurs (CDD) ou ingénieurs	15	23	17	13	11	10
Chercheurs associés				1	2	2
Doctorants	19	14	12	15	14	12
Post-docs	6	7	5	3	2	0

Les moyens attribués à l'équipe se divisent en financement institutionnel (assuré à part égale par le CNRS et Telecom SudParis) et les ressources propres sur contrats. L'équipe fait un effort particulier pour abonder ses ressources propres par des financements complémentaires en ligne avec ses activités de recherche.

Ainsi, la dotation institutionnelle est principalement dédiée à l'incitation des doctorants à publier dans des conférences et journaux de premier rang en prenant en charge les frais de publication et la participation aux évènements. Elle varie autour de 7000€ par an.

L'abondement grâce aux ressources contractuelles permet en plus du financement de thèses, post-docs, ingénieurs, d'aborder le volet applicatif et le prototypage des solutions ainsi que la participation active aux instances de normalisation avec une forte implication et plusieurs contributions aux standards ISO MPEG. Le chiffre d'affaires de l'équipe sur les ressources contractuelles s'élève à 1,26 M€ en 2016 et à 1,87 M€ en 2017. L'équipe a aussi bénéficié de 3 bourses de thèse CIFRE, de 2 bourses de thèse de l'Institut Mines-Telecom (Futur et Rupture) et des ressources d'un projet CARNOT.

### 4.1.3 Profils d'activité

L'équipe ARMEDIA s'appuie sur les méthodes mathématiques et statistiques d'analyse et de modélisation des contenus numériques pour traiter diverses applications. Les thématiques identifiées à sa création et alimentées par diverses actions décrites dans la partie bilan sont structurées selon trois axes :

1. **Image, son et vision** , décliné selon quatre volets : (a) *Santé et assistance à la personne dépendante*, (b) *Identification biométrique*, (c) *Vision par ordinateur et interprétation de scènes*, et (d) *Analyse et simulation de gestes et expressions*;
2. **Modélisation, échange et enrichissement des contenus** , adressant les thèmes : (a) *Modélisation, échange, protection, transmission et adaptation des contenus*, (b) *Indexation et fouille de données multimédia*, (c) *Modélisation des représentations et interactions en réalité augmentée*;
3. **Interaction et interfaces adaptées**, concernant : (a) *Assistant vocal pour personnes dépendantes*, (b) *Robotique d'assistance*, et (c) *Modèles statistiques de dégradation du signal écrit*.

De par le profil appliqué des recherches menées dans l'équipe, l'activité de valorisation et de transfert est très liée à la production scientifique. Elle est perçue comme moteur de production de la connaissance en réponse à des défis socio-économiques actuels. L'appui à la communauté se décline à travers l'implication dans diverses instances de soutien pour la recherche. Si on tente une répartition des efforts selon ces trois axes, nous pouvons les résumer ainsi (voir Annexe 4 pour le détail) :

- activité scientifique (40%) : 54 articles de revue, 151 publications en conférences avec actes, 7 chapitres d'ouvrages et 34 encadrements de doctorants;
- activité de valorisation et de transfert (40%) : 12 contrats de recherche européens, 9 contrats nationaux, 6 projets d'investissement d'avenir et 9 projets industriels ou avec des fondations, pour un chiffre d'affaires dépassant 1,2 M€ par an.
- appui à la communauté (10%) : participation à 37 comités de programme de conférences, 5 organisations de colloques, 3 comités éditoriaux et nombreuses évaluations d'articles, évaluation régulière de projets de recherche, évaluation HCERES du laboratoire Hubert Curien et responsabilités au sein d'instances d'évaluation de pôles de compétitivité (Cap Digital, System@tic) ou de sociétés savantes (FEDEV, AVIESAN), ainsi qu'une forte implication dans les instances normatives avec représentation de la France (ISO MPEG).
- formation par la recherche (10%) : participation à des masters de l'Université Paris-Saclay et de l'Université Paris-Est Créteil, implications dans les activités orientées "Recherche" dans les Voies d'Approfondissement du cursus ingénieur à Telecom SudParis, encadrement de 33 doctorants et de nombreux ingénieurs recherche et stagiaires, création d'un MOOC.

## 4.2 Produits de la recherche et activités de la recherche de l'équipe ARMEDIA

### 4.2.1 Bilan scientifique

Le bilan scientifique de l'équipe sera décliné par la suite conformément aux axes thématiques mentionnés précédemment. A noter que certains thèmes en continuité avec l'activité des membres ayant rejoint le laboratoire à la création de l'équipe font parfois référence à des publications antérieures à 2016 (considérées plus significatives). La distinction de ces publications sera faite de manière explicite dans l'Annexe 4.

## 1. Image, son et vision

**(a) Santé et assistance à la personne dépendante** Dans ce contexte, nos efforts concernent l'imagerie bio-médicale, le son et la télévigilance médicale. Sur le premier volet, ils se dirigent notamment vers l'achèvement de solutions de diagnostic quantitatif assisté par ordinateur (DqAO) applicables cliniquement grâce à des collaborations de plus en plus nombreuses avec des partenaires médicaux. Au niveau méthodologique, nous nous sommes intéressés à des approches fondées sur la morphologie mathématique (en particulier les filtres connexes et les opérateurs géodésiques), la géométrie discrète, les modèles déformables, les ensembles de niveau et la logique floue, et plus récemment, les techniques d'apprentissage profond. Les axes méthodologiques développés et les applications cliniques associées sont déclinés par la suite.

### **(a.1) Marqueurs image pour l'analyse et le suivi des pathologies obstructives de l'arbre trachéo-bronchique** (C. Fetita)

Ayant comme visée clinique l'asthme et les bronchopathies chroniques obstructives (BPCO), l'objectif consiste à fournir au clinicien des outils d'analyse et suivi (qualitatifs, quantitatifs et non-invasifs) du remodelage bronchique induit par une pathologie. A terme, l'objectif ciblé est le phénotypage de l'asthme et des BPCO ainsi que la prédiction de la réponse thérapeutique attendue. Les développements méthodologiques associés ont conduit à des approches automatiques de segmentation 3D de l'arbre trachéobronchique (lumière et paroi [13.J.ar]), d'extraction de l'axe médian pour la navigation endoluminale et la quantification à haut-débit des paramètres bronchiques en section transverse, de caractérisation 3D de la forme luminale et de détection automatique d'anomalies (sténoses, bronchéctasies), de quantification 3D de l'épaisseur pariétale et de son remodelage pour un suivi quantitatif [26.C.ar], ainsi que de modélisation in-silico des voies respiratoires, spécifique au patient, permettant des simulations numériques d'écoulement pour des études fonctionnelles [13.C.ar]. Ce thème a été nourri par plusieurs collaborations académiques et industrielles passées et en cours (recherche et études cliniques) dans un cadre institutionnel, national ou international (projets FP7-AirProm, ANR-OxHelease).

### **(a.2) Analyse des textures pulmonaires en imagerie TDM** (C. Fetita)

Le premier objectif clinique visé est la détection automatique précoce et le suivi quantitatif de l'emphysème et des pneumopathies interstitielles diffuses (PID), pathologies responsables d'une destruction irréversible du parenchyme pulmonaire, à partir de données image acquises en TDM volumique. Les recherches orientées vers des approches combinant filtrage morphologique multirésolution, analyse de graphes et logique floue ont abouti à un outil DqAO automatique avec des résultats très prometteurs en termes de segmentation 3D et classification des pathologies ciblées, pour un suivi quantitatif de l'évolution de la maladie. Des travaux en cours exploitent les nouvelles techniques en apprentissage profond (deep learning) par réseaux neuronaux convolutionnels pour atteindre le niveau de précision requis en pratique clinique [1.C.ar]. Les partenaires cliniques et académiques associés à cette étude incluent l'AP-HP, UPMC et l'Université Paris 13. Ces travaux concernent la thèse de doctorat de Sébastien Tarando, soutenue en 2018.

### **(a.3) Imagerie vasculaire** (C. Fetita)

Deux projets cliniques ont été démarrés sur l'analyse du réseau vasculaire intrapulmonaire en imagerie TDM. Le premier, conduit avec l'Hôpital Bichat, Paris, est fondé sur l'analyse des calibres vasculaires et vise l'étude du remodelage des petits vaisseaux en association avec une pathologie de type BPCO ou AATD (déficience en alpha-1 antitrypsine). Les premiers résultats ont confirmé ceux d'autres études sur la BPCO et ont apporté de nouveaux éléments dans le cas de l'AATD [21.C.ar]. Dans une deuxième investigation initiée en 2016 avec l'Hôpital Ambroise-Paré, Boulogne-Billancourt, les travaux en modélisation du calibre des structures ont été appliquées pour la détection des fistules pulmonaires (pathologies de type aneurisme).

Dans un cadre différent, portant sur l'étude de nouvelles solutions diagnostic couplant des données d'imagerie (micro-TDM, spectroscopie infra-rouge, histologie), biologiques et moléculaires, des études sur l'angiogénèse et le développement tumoral chez la souris sont entreprises dans le cadre du projet PIA IMPACTumors (thèse de Florence Kouvahe).

### **(a.4) Analyse du signal EEG et de la parole pour la détection de maladies neurodégénératives** (N. Houmani, D. Petrovska)

Ce thème vise deux objectifs cliniques : le premier, porté en collaboration avec l'hôpital Charles-Foix (AP-HP), est de détecter la maladie d'Alzheimer et de la discriminer des autres pathologies au stade précoce, par analyse de signaux d'électroencéphalographie (EEG) [12.J.ar]. Le deuxième, en collaboration avec l'hôpital Pitié-Salpêtrière, Paris, exploite la voix et des corrélats d'imagerie chez des sujets à risque de la maladie de Parkinson, dans un but d'aide au diagnostique et pronostique. A ce titre, l'approche originale mise en oeuvre repose sur des enregistrements

vocaux de sujets sains, malades, et à risque, pas seulement dans le milieu hospitalier, mais aussi à distance par téléphone [9.C.ar].

**(a.5) Télévigilance médicale** (N. Houmani, B. Dorizzi, J. Boudy)

Les activités de recherche en télévigilance médicale sont centrées essentiellement sur trois axes (recouvrant en partie le thème 3a) et concernent : (1) la prévention et la détection de chute, ceci par des approches de combinaison/fusion de données/signaux issus de différents capteurs mobiles (portés par la personne) et fixes (environnement, smart home) [14.J.ar], (2) l'analyse du son dans un contexte de surveillance/assistance pour la détection d'évènements acoustiques anormaux, et (3) l'analyse de l'actimétrie ou des mouvements de la personne, de sa localisation [27.C.ar], et de sa qualité du sommeil basée sur des signaux de capteurs thermiques de présence en collaboration avec la société LEGRAND et l'UTC de Compiègne [19.C.ar].

D'une manière plus générale les thèmes suivants sont orientés vers le traitement des données multimédias, décliné selon trois axes méthodologiques.

**(b) Identification biométrique** (B. Dorizzi, D. Petrovska, N. Houmani, S. Garcia, M. El Yacoubi)

Un des défis posés à la reconnaissance par l'iris réside dans la précision de la segmentation de la zone de l'iris dans les images des yeux qui peuvent être de qualité variable, voir médiocre si on relâche les contraintes sur les conditions d'acquisition. Nous nous sommes attaqués à ce problème au cours de trois travaux de thèse. Dans le cadre d'une collaboration avec Thalès, nous avons exploré des modèles ellipsoïdes du contour de l'iris pour améliorer les approximations généralement basées sur des cercles. Dans un autre travail de thèse, nous avons proposé des techniques de super-résolution pour pallier le manque de qualité observé lors d'acquisition à distance de séquences vidéos d'iris. L'originalité de l'approche est d'introduire explicitement une mesure de qualité dans le schéma de fusion [32.C.ar]. Finalement nous avons aussi exploré l'apport d'une segmentation statistique non supervisée de la région d'iris à base de modèles markoviens comme complément des approches de contour usuellement employées [7.J.ar].

Dans la thématique de la biométrie multimodale, nous nous sommes concentrés sur l'exploitation des modalités voix et visage, pour améliorer les performances ainsi que pour s'adapter aux capteurs existants et permettre leur intégration sur des dispositifs mobiles (téléphones portables, tablettes,...) [14.C.ar]. Un autre axe consiste à fusionner les biométries iris et visage pour permettre la création de clefs crypto-biométriques. Celles-ci peuvent remplacer les clefs cryptographiques classiques, avec l'avantage supplémentaires qu'elles fournissent une preuve biométrique de l'identité de l'utilisateur et n'ont pas besoin d'être stockées. Ces résultats innovants prouvent qu'il est possible de générer des clefs cryptographiques suffisamment longues et toujours identiques à partir des données biométriques qui sont variables. En plus, cette méthode a été adaptée pour permettre de révoquer ces clefs crypto-biométriques en cas de compromission. Cette révocabilité assure aussi que notre vie privée est protégée, car les informations biométriques stockées sont différentes pour différentes applications [2.L.ar].

Dans l'analyse de la biométrie comportementale de la signature manuscrite en-ligne, nos travaux ont porté sur une modélisation originale du tracé par un codage d'information par "minimum jerk" classiquement utilisé en analyse du geste [10.J.ar]. Une autre étude sur tablette graphique a consisté à proposer des mesures de qualité du signal pour générer automatiquement des catégories de personnes, permettant de prédire les performances des systèmes de vérification d'identité selon les catégories [8.J.ar], [25.C.ar]. Nous avons également exploité ces critères de qualité pour sélectionner les signatures de référence [25.C.ar] pour la vérification d'identité par la signature en-ligne.

Un autre thème abordé est celui de la biométrie des veines du doigt, où nous avons proposé des techniques originales par deep learning pour l'estimation de la qualité des images, leur segmentation et la génération de veines manquantes dans les images [5.J.ar], sans nécessiter une annotation manuelle pour ces trois tâches. En réidentification de personnes, dans le cadre de la vidéosurveillance, nous avons proposé une nouvelle approche par représentation parcimonieuse qui encode conjointement les paramètres spatiaux et de mouvement [6.J.ar].

**(c) Interprétation de scènes en temps réel** (T. Zaharia)

La problématique bien connue de l'interprétation de scènes à partir de flux vidéos renvoie, aujourd'hui encore et après des décennies de recherche intensive, à des défis majeurs qui restent à être relevés. Dans ce cadre, nous nous sommes intéressés à un contexte d'application spécifique, à savoir l'interprétation en temps réel de scènes urbaines pour construire des systèmes embarqués d'aide à la mobilité des personnes aveugles et malvoyantes.

Les principales contributions apportées s'inscrivent dans le cadre du projet AAL ALICE et concernent [4.J.ar], [3.C.ar] : (1) une méthode de détection en temps réel (15 trames/secondes sur un smartphone Android de type Samsung S6) d'obstacles, à l'aide de techniques d'analyse de mouvement par échantillonnage semi-dense et renfor-

cement de la cohérence spatiale ; (2) une approche de reconnaissance d'obstacles selon une taxonomie simplifiée, comportant mobilier urbain, véhicules, piétons, vélos ..., à l'aide de techniques de classification supervisée, (3) une technique de détection, reconnaissance et suivi d'objets d'intérêt indoor, spécifiés par l'utilisateur, (4) une approche de reconnaissance de bâtiments urbains (landmarks), qui vise à améliorer les capacités de localisation en l'absence de signal GPS, (5) un ensemble complet, intégrant logiciel et hardware sous forme de ceinture connectée, capteurs vidéos et ultrasoniques, ainsi que de dispositifs de rendu audio non-intrusifs.

#### **(d) Modélisation/interprétation d'émotions, intentions, comportements** (T. Zaharia)

Initié en 1995 avec les travaux pionniers de Rosalind Picard au MIT, le calcul affectif est devenu aujourd'hui une branche à part entière des sciences et technologies de l'information. Aujourd'hui, les recherches portant sur le geste manquent cruellement de modèles mathématiques unifiés et génériques.

Dans ce cadre, notre contribution est plurielle. En premier lieu, nous avons repris les concepts développés par le chorégraphe Rudolf Laban pour l'analyse et l'enseignement de la danse classique contemporaine (Laban Movement Analysis : LMA), et proposé leur extension afin d'élaborer un modèle descriptif et générique du geste, fondé sur ses éléments expressifs. Deux différentes approches de reconnaissance de geste ont été ensuite élaborées. Une première utilise le modèle de Laban pour définir un vecteur descripteur global. Elle est validée sur deux expériences de classification faisant intervenir des méthodes d'apprentissage supervisé : la première a trait à la discrimination des actions d'un corpus de référence ; la deuxième vise à reconnaître les émotions musicales que portent les gestes des chefs d'orchestre. Dans une seconde approche dite "dynamique" ou "locale", les caractéristiques du modèle sont reprises pour constituer des descripteurs de trame gestuelle (e.g. pour chaque instant du geste). Une méthode d'affectation douce, couplée avec une procédure d'apprentissage par chaînes de Markov cachés permettent ensuite d'obtenir à tout instant une représentation simplifiée du geste, utilisable pour reconnaître des actions à la volée [3.J.ar], [15.C.ar]. Ces développements se sont inscrits dans le cadre du projet ITEA2 Empathic Products et ont fait l'objet de la thèse d'Arthur Truong (soutenue en septembre 2016).

## **2. Modélisation, échange et enrichissement des contenus**

Les thèmes de recherche développés ici visent plusieurs enjeux technologiques : modélisation, traçabilité, recherche, échange et enrichissement de contenus, et seront abordés selon les axes suivants.

#### **(a) Modélisation de sources naturelles d'information** (M. Mitrea)

La modélisation mathématique des sources naturelles d'information (langage écrit, image, vidéo, ...) est un problème récurrent qui doit suivre, à chaque instant, les évolutions sociétales. Dans ce cadre, nous avons tout d'abord étudié les synergies et les antinomies entre les aspects statistiques et sémantiques (dans le sens d'un observateur humain) d'un même contenu visuel. A cette fin, nous avons considéré deux directions. La première porte sur la possibilité de lier la saillance visuelle (intrinsèquement gouvernée par les mécanismes bio-cognitifs inclus dans le système visuel humain) aux éléments syntaxiques des flux vidéo compressés. L'état de l'art est jalonné par des publications permettant de calculer des cartes de saillance à partir d'un contenu visuel représenté par des pixels. Nos principales contributions consistent en : (1) avancer un cadre méthodologique compréhensif permettant d'identifier, selon leurs comportements statistiques, les éléments syntaxiques MPEG-4 AVC et HEVC avec une représentativité sémantique visuelle ; (2) établir une procédure de test générale, qui permet d'interpréter quantitativement les résultats obtenus et (3) valider ce cadre méthodologique dans une application de protection du contenu visuel apparentée au tatouage robuste. Dans leur ensemble, outre la nouveauté méthodologique, nos résultats ([2.J.ar],[12.C.ar],[11.C.ar]) présentent des gains de 60% à 160% dans la précision et entre 17% et 20% en discriminance, tout en comportant une complexité linéaire de calcul (thèse de Marwa Ammar soutenue juin 2017). Les extensions des outils images vers les signaux audio sont étudiés en [7.C.ar]. La deuxième direction de recherche est l'analyse de l'impact des étiquettes sémantiques normatives (ITU) dans les évaluations subjectives de qualité visuelle. Si l'existence d'une telle influence était déjà suggérée dans l'état de l'art, notre contribution est de définir une approche statistique originale, combinant transformations non-linéaires de variables aléatoires, estimations de loi de probabilité et tests de probabilité et permettant à la fois de prouver l'existence de cette influence et sa quantification, sous la forme d'un coefficient dépendant du type de contenu - vidéo stéréoscopique ou 3D (thèse de Rania Bensaïed soutenue juillet 2018).

#### **(b) Traçabilité des contenus** (M. Mitrea)

Les cadres méthodologiques du fingerprinting et du watermarking se positionnent comme deux repères complémentaires pour assurer le traçage du contenu visuel transféré d'une façon légale. Les techniques de fingerprinting permettent d'identifier tout contenu vidéo à l'aide d'une signature de taille réduite, calculée à partir de la séquence d'origine et stockée dans une base de données. A noter que cette signature doit être invariante aux transforma-

tions du contenu vidéo, qu'elles soient malveillantes ou non (compression, changement de couleurs, échantillonnage spatio-temporel,...). Malgré ce foisonnement strictement technique, de nombreuses questions fondamentales restent posées. En effet, la définition même d'un cadre théorique dédié, unitaire et précis pour les méthodes de fingerprinting reste un desideratum. Cet état de l'art conduit à l'impossibilité de définir le caractère biunivoque, stable et optimal de la signature. Corrélativement, cela impose de vérifier les propriétés de robustesse par critères strictement empiriques, en l'absence d'un modèle d'estimation a priori. Dans ce contexte, nos contributions sont : (1) avancer un modèle théorique pour le fingerprinting, basé sur la théorie de l'information, (2) l'utilisation de ce modèle théorique pour extraire un fingerprint en concordance avec les hypothèses de stationarité et ergodicité; (3) définir une règle statistique pour l'appariement des fingerprints, avec le contrôle des erreurs; (4) dériver l'algorithme scalable sous-jacent qui a permis, pour la première fois, une robustesse aux attaques aux enregistrements par camera externe [9.J.ar].

Sur le volet watermarking, nos travaux nous ont permis à la fois d'élargir le socle théorique antérieurement défini pour prendre en compte le cas du contenu vidéo stéréoscopique et établir la preuve de concepts pour le tatouage semi-fragile du flux compressé MPEG-4 AVC.

Concernant la protection du contenu vidéo stéréoscopique, l'état de l'art est donné par des approches de type 2D+t vers 3D. Dans ce contexte, nos contributions se positionnent en rupture et comportent : (1) la définition d'une nouvelle carte de disparité, établissant synergies entre les particularités du système visuel humain et le comportement non-homogène du contenu visuel stéréoscopique, (2) la spécification d'un cadre méthodologique permettant des complémentarités applicatives entre les paradigmes du tatouage par spectre étalé, du tatouage par information de bord et des simulateurs de type Monte-Carlo pour les transformations du contenu vidéo. Concernant le tatouage semi-fragile des séquences vidéo, l'état de l'art faisait foi d'une large panoplie des méthodes pour les images fixes et/ou les séquences vidéo représentées par des pixels, tout en ignorant la difficulté d'insérer la marque directement dans le flux compressé. Nos contributions consistent en : (1) démontrer la règle optimale d'insertion multi-symbol mQIM (et accroître implicitement la quantité d'information insérée par un facteur  $\log_2 m$ ) dans le sens de la minimisation de la probabilité d'erreur pour une puissance de bruit donnée; (2) spécifier la règle optimale d'extraction de l'information d'auto-authentification des séquences vidéo, dans le sens de la théorie de l'information, (3) formaliser dans un cadre algébrique le problème du drift spatial induit par le tatouage du flux MPEG-4 AVC et résoudre l'équation sous-jacente sous contrainte d'élimination du drift [29.C.ar].

**(c) Fouille de données hétérogènes, distribuées et dynamiques** (T. Zaharia, M. Preda, D. Petrovska, G. Chollet)

Dans le domaine de l'indexation multimédia et de la fouille de données hétérogènes, les travaux réalisés se sont inscrits en continuité naturelle avec les lignes de recherche développées précédemment.

Un premier axe concerne l'indexation par objet vidéo. Les résultats obtenus montrent la pertinence des approches par extension de requêtes, pour résoudre ce problème complexe et mettre au point des procédures de requête partielle fiables et efficaces. Il s'agit dans ce cadre de mettre en oeuvre une méthode d'analyse de graphes de similarités visuelles pouvant identifier des clusters d'apparences pertinents et éliminer en même temps les outliers.

Les méthodes d'analyse de graphe par extension de requêtes ont également été re-considérées dans un contexte différent, qui concerne notamment l'analyse textuelle des sous-titres associés aux programmes de télévision. Dans ce cadre, les techniques de structuration temporelle en plans/scènes/chapitres par attributs visuels ont été hybridées avec de nouvelles méthodes d'analyse purement textuelle par graphes de concepts dominants afin d'obtenir une caractérisation pertinente de contenus multimédias enrichis [5.C.ar].

Dans un contexte plus applicatif, lié aux usages émergents de la ville numérique intelligente (cf. projet Smart-City+), un moteur de recherche dédié pour les services d'hyper-proximité a été développé. Cela concerne notamment l'adaptation de l'ontologie générique WOLF à des objectifs de recherche ciblés.

Enfin, une dernière contribution concerne la mise en place d'une méthode d'extraction d'une signature locale et d'un détecteur de régions d'intérêt invariant aux déformations perspective et aux bruits sur la base de la transformée de Fourier-Mellin quaternionique, ainsi que d'un service web de reconnaissance d'images exploitant cette méthode [8.C.ar]. Cette transformation a été ensuite utilisée dans un schéma de protection de contenu visuel [22.C.ar].

Sur le volet "audio", nos travaux sur le traitement de la parole avec des méthodes complètement "data-driven" ont abouti à une proposition originale d'indexation et de recherche dans des contenus audio, et ceci pour des applications diverses, telles que la diarization du locuteur, la détection de publicité, de morceaux de musique, ... [30.C.ar].

**(d) Réalité augmentée** (M. Preda)

L'approche développée pour la réalité augmentée concerne la mise en place d'un formalisme pour la représentation de la scène 3D des médias, de l'accès aux capteurs [28.C.ar], de l'accès aux ressources déportées, et des interactions utilisateurs [11.J.ar]. Ce formalisme est également la base des outils auteurs qui ont été développés dans une plateforme web, permettant ainsi la facilitation de la création des contenus de réalité augmentée. Suite à des nombreuses contributions dans les instances de normalisation, ce formalisme est actuellement le fondement du nouveau standard MPEG-A : Augmented Reality Application Format.

**(e) Adaptation des contenus pour des échanges sur terminaux légers** (M. Mitrea)

Plusieurs limitations théoriques et méthodologiques ont pu être identifiées dans l'état de l'art pour les représentations des scènes multimédias collaboratives. Tout d'abord, les solutions actuelles n'offrent qu'une émulation des caractères participatif et collaboratif. Or, avec l'augmentation exponentielle des utilisateurs, le problème du passage à l'échelle ou de la scalabilité se pose de façon cruciale et se traduit rapidement par un blocage au niveau du temps de calcul et du trafic réseau. Ensuite, notons qu'il n'y a aucune méthode de compression capable d'assurer pour chaque type de contenu une qualité variable (en fonction de ses particularités perceptuelles) tout en minimisant la complexité de calcul. Enfin, le travail collaboratif impose de disposer de codecs (codage et décodage) temps réel, pour gérer à la volée les flux autoproduits et enrichis, sur des terminaux munis de ressources et de capacités fort différentes (stations, clients légers, téléphones portables...) : il n'existe pas de schéma de compression scalable permettant de résoudre le compromis complexité - qualité - taux de compression. Dans ce contexte, nos contributions peuvent être structurées sur quatre directions : (1) la définition, implantation et validation des mécanismes d'accès partagé (collaboratif et participatif) aux éléments d'une scène multimédia ; (2) l'étude, la spécification et l'implantation des mesures objectives mettant en évidence la qualité d'expérience lors de la perception d'un contenu multimédia ; (3) mise au point d'un ensemble de gestion des schémas de compression permettant de trouver l'équilibre fonctionnel entre taux de compression, qualité visuelle et complexité d'encodage/décodage ; (4) validation de ces concepts dans un espace collaboratif de travail médical grandeur réelle, en collaboration avec quatre hôpitaux européens [20.C.ar].

**3. Interaction et interfaces adaptées**

**(a) Assistant vocal pour des personnes dépendantes** (J. Boudy)

Ce thème traite de la reconnaissance automatique de la parole pour un usage par des personnes dépendantes par adaptation de Modèles de Markov Cachés acoustiques [18.C.ar]. Il s'appuie sur des modèles de langage générique et spécifique pour la gestion du dialogue avec un majordome vocal.

**(b) Interfaces homme-machine** (N. Houmani)

Nos travaux portent sur le développement d'interfaces cerveau-machine pour l'analyse et le suivi de l'attention visuelle soutenue et sélective, par le traitement de signaux EEG. Le but est de mieux comprendre le processus cognitif de l'attention, l'engagement, le désengagement lors d'une tâche ainsi que la stratégie adoptée lors d'une tâche sélective (neurosciences cognitives et comportementales) [6.C.ar]. Nous avons aussi développé des jeux sérieux sur tablettes afin d'étudier l'interaction des personnes avec l'outil numérique et d'évaluer certaines fonctions cognitives comme l'attention, la mémoire et la planification, mais aussi la motricité fine de la main et la coordination oeil-main.

**(c) Modèles statistiques de dégradation du signal écrit** (S. Garcia, N. Houmani, M. El Yacoubi)

Nos travaux portent sur l'analyse de l'écriture manuscrite acquise sur tablette graphique, par des méthodes d'apprentissage statistique. Nous avons étudié l'évolution de l'écriture au cours de l'âge [24.C.ar] et montré que, contrairement à la littérature, les personnes âgées ont plusieurs profils d'écriture vis-à-vis de paramètres spatio-temporels. Afin de caractériser la maladie d'Alzheimer à ses débuts, nous avons proposé une approche non supervisée sur des paramètres semi-globaux et locaux de l'écriture, là où la littérature n'exploite que des paramètres globaux. Nous avons dégagé des groupes divers liés aux classes, puis agrégé via un formalisme bayésien l'information qu'ils portent, pour la classification. Nous avons montré dans [1.J.ar] l'apport significatif de notre approche. Un autre axe de recherche concerne l'étude de l'impact du changement de plateforme sur la qualité de la signature manuscrite en ligne. En effet, notre geste change et s'adapte au support de manière inconsciente. On s'intéresse à la quantification de la dégradation du tracé par des modèles statistiques lorsqu'on passe d'une tablette Wacom, munie d'une feuille de papier, à des écrans tactiles (tablettes et Smartphones) [1.L.ar].

## 4.2.2 Faits marquants de l'équipe ARMEDIA

En raison de l'activité équilibrée de l'équipe sur les trois volets, production scientifique, valorisation et appui à la communauté, nous pouvons considérer un échantillon représentatif des réalisations recensées en Annexe 4 qui vont pour certaines au delà des missions habituelles du chercheur.

1. Organisation d'une session spéciale sur "Visual Saliency and Image/Video Representation Formats" à IPTA 2017 (International Conference on Image Processing Theory, Tools and Applications).

2. Forte structuration de l'activité de recherche en santé et autonomie autour du développement de collaborations pérennes avec des hôpitaux et centres de soin franciliens et de la mise en place de deux plateformes à Telecom SudParis - BioMICA et EVIDENT. On note un fort développement de l'activité sur le quinquennat (thèses, publications, projets financés).

3. Obtention de trois projets H2020 sur la période. Le taux de réussite pour ce type de dépôt est très faible, et ce succès témoigne de la qualité des propositions.

4. Création d'une spin-off, uStartApp (www.inside.cloud). Elle témoigne d'une part de l'implication de l'équipe dans les activités de transfert et valorisation de la recherche et, d'autre part, de la pertinence des résultats obtenus permettant une telle exploitation.

5. Publication de la norme ISO Augmented Reality Application Format (MPEG-A Partie 13) comportant le langage de description des contenus RA interactifs proposé par les membres de l'équipe.

6. Prise de responsabilité au sein d'instances d'évaluation de pôle de compétitivité : vice-présidence du groupe thématique *Ville Numérique* du Pôle System@tic et de la commission thématique *Contenus* du Pôle Cap Digital.

## 4.3 Organisation et vie de l'équipe ARMEDIA

L'animation de l'équipe est assurée par des séminaires réguliers ouverts pour la plupart à l'ensemble du laboratoire. Un binôme appuyé par le responsable d'équipe en assure l'organisation. Du fait de la bilocalisation des membres, les séminaires sont planifiés en mode visioconférence en s'appuyant sur l'infrastructure de Telecom SudParis. Les doctorants de l'équipe sont incités à présenter leur travaux à ces occasions.

Le responsable d'équipe gère le budget interne (dotation institutionnelle) et décide de l'attribution égalitaire des subventions pour la participation aux conférences ou l'achat d'ouvrages, tout en appliquant les critères d'excellence établis par le laboratoire. A noter que deux doctorants de l'équipe ont obtenu le prix du jury et le prix du public lors de la Journée des Doctorants de SAMOVAR organisée en 2017 à Telecom SudParis.

## 4.4 Analyse SWOT de l'équipe ARMEDIA

L'analyse de nos activités, de nos résultats et de notre positionnement conduit à l'auto évaluation suivante :

### Points forts

- Une recherche équilibrée entre activité académique et valorisation.
- Dynamique croissante au niveau des projets collaboratifs.
- Bonne implantation dans les instances d'évaluation des pôles de compétitivité System@tic et Cap Digital.
- Forte représentation au niveau de l'activité de standardisation.
- Couverture thématique large du domaine du multimédia.

### Points faibles

- La couverture thématique large du domaine ne favorise pas l'interaction forte entre les membres de l'équipe.
- L'investissement dans des montages de projets, dans un contexte de faibles taux de succès, peut impacter l'activité de publication.
- La bilocalisation Evry-Saclay limite l'interaction directe entre les chercheurs de l'équipe.

### Opportunités

- Effervescence de la thématique Data IA qui impacte fortement les domaines d'expertise de l'équipe.
- Rapprochement des acteurs du plateau de Saclay avec l'intégration à la Faculté DSIT (Data Science and Information Technology) de NewUni et avec la bilocalisation de Telecom SudParis à l'horizon 2019, apportant des bénéfices au niveau des collaborations et des recrutements en thèse.
- Thématiques de l'équipe positionnées dans les axes stratégiques de l'Institut Mines-Telecom favorisant des



interactions plus larges au niveau du groupe et le soutien institutionnel.

### Menaces

- Forte compétition sur la thématique Data IA augmentant la difficulté de proposer des contributions scientifiques originales
- La bilocalisation de l'équipe présente un risque de clivage même au sein de thématiques de recherche proches dans un contexte compétitif influencé par des partenaires/concurrents locaux forts
- Incertitude sur les transformations et les restructurations à venir au niveau des laboratoires de recherche et des Ecoles Doctorales du plateau de Saclay.

## 4.5 Projet scientifique de l'équipe ARMEDIA

Ancrées dans le domaine multimédia, les activités de l'équipe se verront naturellement impactées par les développements récents en intelligence artificielle (IA), avec une "explosion" des techniques par apprentissage machine dans tous les axes thématiques abordés. Cela favorisera par ailleurs des actions transverses à plusieurs équipes au niveau du laboratoire autour de plusieurs domaines applicatifs tels que *Santé et autonomie*, *Ville intelligente*, *Industrie du Futur* et *Internet des Objets*.

Le projet de l'équipe s'articule autour de deux défis scientifiques déclinés selon les trois axes thématiques identifiés dans la partie bilan et conduisant à diverses applications.

### DÉFI 1. APPROCHES INTÉGRATIVES POUR DES DONNÉES HÉTÉROGÈNES MULTICAPTEUR

Ce défi concerne le traitement conjoint de données hétérogènes et multi-échelles provenant de plusieurs capteurs (ou sources d'information), s'appuyant sur des démarches pluridisciplinaires. Les domaines applicatifs visés sont la santé, notamment le volet médecine personnalisée, et l'industrie du futur. Il s'agit non seulement de disposer d'une information variée et complémentaire (son, imagerie 2D, 3D, 3D+t multimodale, vidéo, signaux, données cliniques, démographiques, génomiques, etc.) mais également de l'analyser de manière conjointe en intégrant des modèles de raisonnement fondés sur des ontologies, la fusion de données, les approches statistiques, géométriques et les graphes, tout en s'appuyant sur les nouveaux paradigmes de l'intelligence artificielle. L'objectif est d'aboutir à une analyse discriminante de ces données (détection/reconnaissance, classification) pour une aide à la décision, avec un pouvoir prédictif relatif au suivi d'un traitement (médecine personnalisée) ou au suivi d'une activité dans un flux (multimédia). Le positionnement de ce profil de recherche par rapport à la stratégie CNRS est en ligne avec le défi Imag'In. Les axes thématiques associés sur le volet applicatif sont *Image, son et vision* et *Interaction et interfaces adaptées*, déclinés selon les actions suivantes.

#### (a) Imagerie biomédicale (C. Fetita)

Sur le volet imagerie médicale, nous allons poursuivre les travaux sur la détection et caractérisation des pathologies interstitielles diffuses pulmonaires et sur l'imagerie vasculaire. Nous nous intéressons à l'extension des concepts des réseaux neuronaux convolutifs par intégration des outils de morphologie mathématique. Ainsi nous visons le renforcement des "features" extraites avec des couches convolutives par des filtrages non-linéaires au moyen d'opérateurs morphologiques (opérateurs connexes, en particulier).

Le thème *bioimagerie* verra la naissance d'un projet transverse au laboratoire (projet CellAvIe) en collaboration avec le Genethon sur l'analyse et le suivi sans marquage de cellules (en particulier les cellules souches hématopoïétiques). Le projet repose sur le développement d'une technique de microscopie à grand champ par ptychographie de Fourier (Y. Gottesman - équipe TIPIC), des techniques d'analyse image par IA et calcul parallèle pour la détection, (E. Brunet - équipe ACMES, C. Fetita - équipe ARMEDIA), le suivi des cellules (D. Clark - équipe TIPIC) et leur caractérisation en termes de différenciation cellulaire (D. Stockholm - Genethon). Le projet est hébergé sur la plateforme BioMICA de Telecom SudParis (intégrée au cluster Genopole) et a obtenu un financement de Genopole pour la partie équipement.

#### (b) Analyse des signaux physiologiques et assistance à la personne dépendante (N. Houmani, J. Boudy, D. Petrovska, G. Chollet)

L'activité en analyse de signaux d'électroencéphalographie (EEG) cliniques se poursuit vers la prédiction de profils de personnes à risque de la maladie d'Alzheimer et la détection au stade précoce de troubles neuro-dégénératifs en collaboration avec l'hôpital Charles Foix et La Pitié Salpêtrière (APHP). Elle est complétée par

une deuxième activité sur l'analyse et le suivi de l'attention visuelle par l'EEG et les mouvements oculaires, visant à détecter des détériorations pouvant refléter une fatigue agüe ou une pathologie naissante.

Nous allons poursuivre notre activité sur l'analyse de la voix pour la détection précoce de la maladie de Parkinson et des corrélats avec l'imagerie (IRM). Cette direction sera complétée par l'analyse des signaux vocaux acquis sur des lignes téléphoniques qui permettront un suivi à distance, sans nécessité de déplacement des patients à l'hôpital. Un tout nouvel axe sera dédié à l'analyse des voix normales mais affectées par les émotions.

Concernant l'étude des modèles de dégradation du signal écrit, on vise à étendre l'analyse de signaux manuscrits à des productions graphiques non contraintes, comme la signature manuscrite qui est un processus balistique, le dessin, des tracés manuscrits issus de jeux sérieux sur tablettes (trajectoires) ou de tests neuropsychologiques. Ces recherches ont deux cibles distinctes : (1) identifier des marqueurs prédictifs de maladies neurodégénératives ; (2) caractériser les troubles neuro-développementaux pour leur détection très précoce, lors de la petite enfance (déficit attentionnel, dyspraxie, dyslexie, dysgraphie, autisme). Ici, l'analyse de gestes moteurs fins sera couplée à celle d'autres signaux séquentiels issus de divers capteurs (oculomètre, Leap Motion). L'objectif est de détecter le déficit, et de concevoir des rééducations ciblées (visuelles, gestuelles, graphiques), via des jeux sérieux ou des interfaces ludiques adaptées à la personne.

L'activité sur le développement de "jeux sérieux" sur PC et écrans tactiles se poursuit pour la prévention, la détection et la réhabilitation, en particulier la remédiation cognitive traitée par le neurofeedback EEG. Concernant l'analyse comportementale, nous allons nous concentrer sur le profilage de personnes, à travers leur signature manuscrite et leurs interactions avec les objets tactiles.

Nos travaux sur la télévigilance médicale se poursuivent sur deux volets. Le premier vise la détection et prévention des chutes par l'analyse de la démarche en cours de fragilisation, ceci par recours à des techniques de traitement du signal, de reconnaissance des formes et de fusion. L'usage tout récent du tapis actimétrique ou sol sensitif sur lequel les investigations ont déjà commencé (projet Fonds d'amorçage SolSens, 2017) constitue un axe nouveau et prometteur. Le deuxième thème traite du suivi des activités quotidiennes (ADL) en environnement domestique pour la prévention de stades progressifs de fragilisation de la personne âgée seule. Ces travaux sont en relation avec le premier volet, notamment à travers l'usage de capteurs de présence et de son, et l'analyse des signaux correspondants. En parallèle, l'analyse de la démarche par des capteurs actimétriques sera appliquée à la détection de troubles moteurs et l'évaluation de la charge cognitive de la personne.

**(c) Identification biométrique et sécurité** (N. Houmani, S. Garcia, D. Petrovska)

Dans l'analyse de la biométrie comportementale de la signature manuscrite en-ligne, un volet de recherche porte sur l'analyse de traces en criminalistique ayant comme objectif l'identification du scripteur et la restitution d'un contexte par analyse 3D de l'environnement de la trace, visant l'aide à la décision pour l'expert. Nos travaux sur la biométrie audio-visuelle se poursuivront en exploitant des nouveaux paradigmes de "Deep Learning", notamment dans le cadre de la biométrie révocable, pour assurer la protection de la vie privée de la personne.

**(d) Reconnaissance de gestes et suivi d'activités dans les scènes** (M. El Yacoubi, M. Preda)

Un de nos volets de recherche sera centré autour de la modélisation du comportement humain avec des méthodes de machine learning. Le champ d'applications est assez large, allant du mouvement micro, tel que l'analyse de l'écriture manuscrite en e-santé ou la reconnaissance des gestes de la main, au mouvement macro tel que l'analyse de la mobilité humaine dans le contexte des villes intelligentes. Dans chaque cas, la complexité du problème et la qualité de données guideront le choix de modèle, supervisé ou non, à développer. Les travaux entamés dans le cadre du projet H2020 SILENCE sur la reconnaissance des mouvements et gestes à l'aide des capteurs ultrason vont se poursuivre par la mise en place d'une architecture de classification 3D CNN. Les investigations porteront sur le type des données, la création de la base d'apprentissage et l'optimisation de l'architecture. En collaboration avec un partenaire industriel (projet PIA IRONMAN), le volet suivi d'objets va se poursuivre par des expérimentations en milieu réel de la production industrielle où les difficultés sont dues à des aléas sur les conditions extérieures (mouvements, occlusions, hétérogénéité des environnements). Les approches par apprentissage seront évaluées et combinées à la reconstruction géométrique partielle des objets et scènes.

**(e) Analyse de scène pour contenus TV diffusés** (T. Zaharia)

Ces développements concernent la détection et la reconnaissance de visages, la détection de texte, ou encore l'identification des visages parlants à partir de contenus TV diffusés. Le coeur méthodologique des approches qui seront développées s'appuie sur des réseaux de neurones profonds, les enjeux se situant aux niveaux de la personnalisation des catégories à détecter, ainsi que de l'optimisation algorithmique, indispensable lorsqu'on souhaite traiter des contenus en streaming. Enfin, un dernier volet méthodologique concerne la synchronisation des sous-

titres par leur mise en correspondance/alignement avec le texte extrait de la bande sonore, à l'aide des techniques speech-to-text. Ces développements s'inscrivent dans le cadre du projet PIA SUBTIL, piloté par France Télévisions.

## DÉFI 2. ANALYSE ET TRAITEMENT DE DONNÉES MULTIMÉDIAS MASSIVES CONNECTÉES

Ce défi vient en support du deuxième axe de recherche de l'équipe - *Modélisation, échange et enrichissement des contenus* - dans un contexte applicatif lié aux nouvelles technologies de l'industrie du futur et de la ville intelligente, régi par l'expansion continue des masses de données générées entre autres par les nouveaux dispositifs connectés. Lorsque ces dispositifs comportent des capacités multimédias, avec une partie de traitement embarqué des données, on parle de "internet of media things". L'exploitation de ce type de données massives et hétérogènes renvoie vers des problématiques d'analyse temps réel en environnement distribué/embarqué pour la reconstruction de scènes, la visualisation et l'interaction, l'extraction de connaissances et l'apprentissage, la confidentialité et la sécurité des contenus. Un passage à l'échelle "objets multimédias connectés" soulève le besoin d'approches nouvelles, s'appuyant sur l'expertise existante, pour développer les solutions d'analyse en environnement multimédia connecté temps réel. Le positionnement de ce profil de recherche par rapport à la stratégie CNRS est en ligne avec le défi MASTODONS. Sur le volet applicatif, cet axe répond aux besoins divers soulevés par l'industrie du futur et la ville numérique (mobilité, réseaux de distribution, centres de soin, urbanisme et modélisation 3D, surveillance, sécurité, résilience,...) et se décline selon les actions suivantes.

### (a) Modélisation de sources naturelles d'information (M. Mitrea)

La modélisation mathématique des sources naturelles d'information est un problème récurrent qui doit suivre, à chaque instant, les évolutions sociétales. Ces travaux se poursuivront avec la modélisation des codes logiciels (codes source ou exécutable) qu'un utilisateur malveillant insère dans un autre contenu (comme les séquences vidéo, par exemple). Dans ce cadre, le défi est double : non seulement résoudre un problème de modélisation statistique (établir un modèle général pour une source d'information a priori non-stationnaire et dépendante du contexte) mais également un problème d'identification d'une source très faible (qui émet avec une cadence inférieure par un facteur de 10-24 par rapport au bruit et dont la puissance moyenne de l'échantillon émis est inférieure de 50dB par rapport au bruit).

### (b) Traçabilité des contenus (M. Mitrea)

Dans les années à venir nous ciblons d'établir des synergies théoriques et applicatives entre les techniques de traçage du contenu visuel et les solutions basées sur des blockchains. A noter que les blockchains ne sont plus aujourd'hui ni des solutions miracles pour un enrichissement très rapide ni des boîtes noires à caractère ésotériques. Au contraire, elles concernent des communautés open-source de plus en plus variées et lancent aux chercheurs de vrais défis (innovation au niveau du cryptage, réduction de la complexité de calcul, efficacité énergétique, etc.). Dans ce cadre émulateur, nos activités porteront sur la possibilité de tracer, grâce aux solutions de type blockchain, des contenus visuels différents d'un point de vue représentation numérique mais identiques d'un point de vue perception humaine et/ou sémantique.

### (c) Compression de contenus et traitement distribué (M. Preda, T. Zaharia)

Un premier thème adresse le problème de compression très bas débit d'images fixes pour répondre à un besoin de transmission de ce type de contenu sur des réseaux 2G-SMS, encore largement répandus dans les pays émergents (thèse CIFRE de Sébastien HAMIS avec la société BeBound). Les pistes de développement identifiées concernent l'exploitation des réseaux de neurones profonds pour (1) l'optimisation/l'adaptation des différentes étapes de la chaîne de codage et (2) la mise en oeuvre de nouveaux paradigmes de compression fondés sur des réseaux de type GAN (Generative Adversarial Networks).

Les avancées récentes des technologies de reconstruction 3D à partir des images ou des acquisitions laser ont conduit à l'utilisation de plus en plus répandue des nuages de points 3D. Nous nous intéressons à une modélisation par octree conduisant à obtenir un codec comportant le contrôle fin du nombre des points reconstruits. Des résultats préliminaires ont conduit à l'initialisation d'un projet de standardisation ISO. Les travaux en cours sont orientés sur deux aspects : l'élaboration de nouvelles méthodes pour accélérer la capture (projet industriel SAM) et l'amélioration de l'approche actuelle de compression par projection 2D (projet industriel APP).

L'essor des approches basées sur l'apprentissage profond et le besoin de les utiliser dans un contexte embarqué ou mobile conduit à une nouvelle problématique, la compression des paramètres d'un tel réseau. Nous allons nous intéresser à l'adaptation des techniques de compression de signal pour ce nouveau type de contenu et à la

transmission temps réel de ces paramètres (projet SATT CBORBIG).

Finalement, le volet algorithmique distribué dans le cloud va se poursuivre par des expérimentations avec des partenaires industriels (projet EIT Digital AR4Omnisales, et projets industriels BEAR et Actiluna) afin de rendre les approches plus robustes à la variété des contenus et usages. Ce volet sera complété par la mise en place des interfaces de communication pour les objets connectés multimédia (MThings) dans le cadre du développement du standard ISO MPEG IoMT auquel l'équipe participe activement.

**(d) Adaptation des contenus pour des échanges sur terminaux légers nomades** (M. Mitrea)

Dans les années à venir nous allons tenter de concevoir et évaluer une couche supérieure de virtualisation, non plus orientée vers les données et leurs statistique, comme nous l'avons fait dans les années passés, mais vers les aspects sémantiques et cognitifs. D'une façon pragmatique, les travaux vont commencer par l'utilisation des techniques de type IA pour identifier si une telle approche est réaliste ou non et se poursuivra par des itérations successives afin d'aboutir éventuellement à un modèle sémantique pour les échanges entre les objets connectés et les autres ressources de calcul et/ou stockage.

**(e) Analyse, décision et optimisation de flux d'information par apprentissage** (T. Zaharia)

Ce nouvel axe regroupe des recherches sur les techniques d'apprentissage machine appliquées à des processus décisionnels fondés sur l'analyse de données multimédias massives. Un premier volet est dédié à simuler, visualiser et optimiser le trafic urbain à l'aide de techniques d'apprentissage. Après de premiers résultats prometteurs concernant l'optimisation des cycles de feux urbains, qui ont démontré qu'il est possible de décongestionner le trafic en appliquant des méthodes de type PSO (Particle Swarm Optimization), les travaux futurs porteront sur les méthodes de prédiction (temps d'attente des bus, taux d'occupation des bus,...). Ces travaux sont réalisés dans le cadre de la thèse de Dancho PANOVSKI et sont supportés par le projet FUI ETS (Electronic Ticketing System) qui vise à proposer des solutions de mobilité intelligente pour la ville numérique.

Un deuxième aspect concerne l'apprentissage multimodal actif pour la détection d'objet d'intérêt opérationnel dans les contenus multimédias. Les méthodes de classification par apprentissage profond reposent le plus souvent sur l'abondance des jeux d'entraînement et sur la qualité de leurs annotations. Néanmoins, les apprentissages conduits sur des jeux trop restreints risquent de conduire à des problèmes de sur-apprentissage. Cela représente un vrai verrou technologique pour les applications qui nécessitent de considérer des ensembles de catégories évolutifs et personnalisables, comme c'est le cas pour la détection d'objets d'intérêt opérationnels. Pour relever ce défi, nous allons considérer des techniques d'apprentissage dit actif et semi-supervisé. Une analyse multi-modale (texte, image et audio) sera également considérée pour renforcer le processus de classification et valider les hypothèses. Ces travaux concernent la thèse CIFRE de Paul Guélorget (en coopération avec Airbus).

Le troisième volet s'intéresse à la fouille de données de trajectoires client pour une application d'achat dans des centres commerciaux. Le contexte de ces travaux est celui des applications de "retail" et vise à permettre aux gestionnaires des magasins de piloter et optimiser leurs activités. Cela implique l'élaboration et la mise en oeuvre de mesures de fréquentation, s'appuyant sur le suivi des personnes et de leurs trajectoires dans les magasins. L'objectif est d'élaborer, à partir des trajectoires utilisateur, des techniques d'analyse, de classification et de prédiction pouvant aboutir à différentes applications, incluant la catégorisation des profils utilisateurs, l'identification des corrélations entre les produits achetés, ou encore la prédiction d'un ensemble d'éléments relatifs au processus de vente. Ces travaux s'inscriront dans le cadre de la thèse CIFRE de Safa Ben Youssef, en coopération avec la société Cliris.

# 5 Équipe METHODES

## 5.1 Présentation de l'équipe METHODES

### 5.1.1 Introduction

L'équipe METHODES s'est constituée en 2011 autour de trois groupes thématiques qui ont en commun l'étude de méthodologies et leurs mises en œuvre dans divers domaines d'application : un groupe Optimisation, avec des travaux portant essentiellement sur la recherche opérationnelle et l'optimisation continue et combinatoire, notamment dans les graphes, un groupe Méthodes formelles avec des travaux portant sur la vérification, les preuves et le génie logiciel formel, ainsi que le test et le monitoring, et un groupe Evaluation de performance, se focalisant essentiellement sur des outils tels que les Chaînes de Markov, les files d'attente et la simulation.

L'équipe METHODES s'est enrichie en décembre 2015 avec l'arrivée de Natalia Kushik, et en 2016 avec l'arrivée de sept nouveaux membres, permanents et associés, de l'ENSIIE, notamment dans les groupes Optimisation et Méthodes formelles. Michel Marot a rejoint l'équipe R3S en juin 2016 et Monique Becker s'est installée à NanoInnov, sur le plateau de Saclay en 2013. Jérémie Jakubowicz est mis à disposition depuis janvier 2018.

Durant la période allant de janvier 2013 à juin 2018, l'équipe METHODES a aussi accueilli 12 post-doctorants, 49 doctorants et 8 supports de recherche.

Plusieurs membres de l'équipe METHODES collaborent avec des membres d'autres équipes de Samovar, ce qui donne à l'équipe METHODES un rôle central, au croisement des différents domaines méthodologiques ainsi que de divers contextes d'applications, tels que les systèmes complexes, l'énergie, et les réseaux.

### 5.1.2 Effectifs et moyens de l'équipe

Au 30 juin 2018, l'équipe METHODES compte 15 membres permanents et 4 membres associés. Le tableau 5.1 indique les membres permanents de l'équipe, répertoriés selon les trois groupes thématiques. L'évolution de leur nombre par année est indiquée dans le tableau 5.2.

Les membres associés sont : Alain Faye, Julien Forest, Jérémie Jakubowicz et Massinissa Merabet.

De 2013 à 2018, l'équipe METHODES a accueilli 12 post-doctorants (pour un total de 126 hommes-mois, soit une moyenne de l'ordre de 11 mois par post-doc), 49 doctorants et 8 supports de recherche (61 hommes-mois, soit une moyenne de l'ordre de 8 mois par personne en support).

Le tableau 5.3 retrace l'évolution du nombre des doctorants de 2013 à 2018. Leur financement a été comme suit :

- 20 sur contrats, soit des financements sur contrat de recherche, mais aussi bourses DIGICOSME et CAP-DIGITAL
- 9 CIFRE
- 8 Etranger, ceci désigne quelqu'un payé par une bourse d'un gouvernement étranger, ou par Erasmus, ou salarié à l'étranger
- 6 Institut Mines-Télécom (IMT) - Télécom SudParis (TSP), soit des bourses de l'IMT ou de TSP ou CARNOT, le programme Futur et Rupture de l'IMT, les bourses d'excellence IMT-TSP, etc
- 3 Externes, ceci désigne un salarié en France hors IMT, ou quelqu'un encadré en partie par un enseignant-chercheur de Samovar et inscrit ailleurs en France
- 3 cotutelles

La durée moyenne d'une thèse dans notre équipe est de 40.42 mois.

Groupe	Membres	
Optimisation	Walid Ben-Ameur	
	José Néto	
	Dimitri Watel	
Méthodes formelles	Guillaume Burel	
	Ana Rosa Cavalli (professeure émérite)	
	Catherine Dubois	
	John Paul Gibson	
	Natalia Kushik	
	Stéphane Maag	
	Amel Mammar	
	Renaud Rioboo	
	Evaluation de performance	Tulin Atmaca
		Monique Becker (professeure émérite)
Hind Castel		
	Tijani Chahed	

TABLE 5.1 – Membres permanents de l'équipe METHODES au 30/6/2018

2013	2014	2015	2016	2017	2018
12	12	13	16	16	15

TABLE 5.2 – Nombre des membres permanents de l'équipe METHODES

Nous notons une baisse en nombre de doctorants ces dernières années, due essentiellement à la baisse globale du nombre de contrats, ainsi qu'au nouveau statut comme professeure émérite de Monique Becker (depuis novembre 2013) et Ana Cavalli (depuis novembre 2014) qui ne leur permet plus d'assurer des directions de thèses.

2013	2014	2015	2016	2017	2018
23	25	18	17	18	16

TABLE 5.3 – Nombre des doctorants de l'équipe METHODES

En ce qui concerne les ressources, le tableau 5.4 montre le détail des ressources contractuelles, pour un volume total d'environ 2,8 Meuros sur la période allant de janvier 2013 à décembre 2017.

Nous notons une grande diversité de types de contrats, auxquels s'ajoutent 9 thèses Cifre, pour la même période. Nous notons aussi une baisse globale due à la forte concurrence à l'accès aux financements ces dernières années, même si les montants provenant de l'ANR et PIA résistent bien à cette tendance.

### 5.1.3 Profils d'activité

Les membres de l'équipe METHODES sont très investis dans les quatre profils d'activités suivants :

- les activités scientifiques représentent environ 40% de notre effort total, avec 88 articles dans des revues, dont 41% classés A\*/A selon le référentiel Era Core, et 208 articles dans les actes de congrès internationaux, dont 46% classés A\*/A et B selon le même référentiel.
- les activités de valorisation et de transfert représentent environ 25% de notre effort total, avec une implication importante dans les projets, comme indiqué ci-dessus.
- l'appui à la communauté représente environ 20% de notre effort total, avec une forte présence dans des activités d'organisation de colloques et de conférences, dans les comités de rédaction, et dans les comités d'évaluation de projets et de travaux de recherches (thèses et HDR notamment).
- la formation par la recherche représente environ 15% de notre effort, avec des coordinations et enseignements de modules dans différentes voies de spécialisation à Télécom SudParis et plusieurs Masters à l'université Paris-Saclay entre autres.

	2013	2014	2015	2016	2017
Total	642259	671004	585355	452866	449 866
Internationaux	7000	9620	8880	0	0
Européens	431783	373892	224953	119640	142725
ANR	61726	51406	111480	57688	156381
PIA	0	0	69315	124167	105235
Autres publics	41086	5610	0	0	0
Industriels	93367	185534	136687	127840	44800
Institut Carnot	7297	34942	34040	23530	0
Financement innovation (SATT, BPI, etc)	0	10000	0	0	0

TABLE 5.4 – Moyens financiers sur contrats de l'équipe METHODES

## 5.2 Produits de la recherche et activités de la recherche de l'équipe METHODES

### 5.2.1 Bilan scientifique

Le bilan global en terme de publications est d'environ 296 articles : 30% en revues et 70% en conférences. Le détail année par année est montré dans le tableau 5.5. Notons que les publications pour 2018 sont en cours de saisies dans notre base de données dédiée à cet effet.

	2013	2014	2015	2016	2017	2018 (Juin)	Total	Pourcentage
Revues	10	12	16	19	15	16	88	30 %
Confs	37	36	39	49	36	11	208	70 %

TABLE 5.5 – Publications de l'équipe METHODES

Nous notons que 41% des articles en revues sont classés A\*/A selon le classement Era Core et 46% des articles en conférences sont classés A\*/A et B selon ce même référentiel. Le détail par année se trouve dans le tableau 5.6.

	2013	2014	2015	2016	2017	2018 (Juin)	Total	Pourcentage
Revues A*	2	1	1	5	0	1	10	11 %
Revues A	2	3	3	4	6	8	26	29 %
Revues B	2	2	3	4	3	1	15	17 %
Confs A*	1	0	1	1	0	1	4	0,02 %
Confs A	6	2	4	5	3	0	20	10 %
Confs B	10	10	12	22	13	4	71	34 %

TABLE 5.6 – Classement des publications de l'équipe METHODES

Dans ce qui suit, nous citerons dans le descriptif de notre bilan environ 20% des publications en revues et 20% des publications en conférences, comme demandé dans l'annexe 4. Cette règle implique que certaines sous-sections n'auront pas de références, d'autres auront comme références des papiers dans des conférences alors même que des papiers dans des revues existent pour le même travail. Une annexe à ce rapport contient l'ensemble de nos publications.

### Bilan groupe Optimisation

Les membres depuis 2013 de ce groupe sont Walid Ben Ameer, José Neto, et Jérémie Jakubowicz. Nous comptons à partir de 2016 un nouveau membre permanent, Dimitri Watel, et un nouveau membre associé, Alain Faye, et depuis 2017 un nouveau membre associé, Massinissa Merabet. Jérémie Jakubowicz est membre associé depuis janvier 2018, suite à sa mise à disposition. Les travaux de ce groupe portent essentiellement sur la recherche

opérationnelle et l'optimisation continue et combinatoire, notamment dans les graphes, avec des applications dans les réseaux et l'analyse de données notamment.

### **Graphes et optimisation combinatoire**

Une activité importante du groupe dans cette thématique concerne les problèmes de  $k$ -séparateur. L'idée ici est de caractériser plusieurs classes de graphes pour lesquels le problème se résout polynomialement, en proposant des algorithmes d'approximation, et en développant des approches polyédrales pour le problème. Nous avons aussi travaillé sur des problèmes d'orientation de graphes où nous cherchons à optimiser une fonction qui dépend des déséquilibres des sommets (la différence entre nombre d'arcs entrants et nombre d'arcs sortants). Des approches polyédrales ont été proposées ainsi que des algorithmes d'approximation, des résultats d'inapproximabilité et des caractérisations de certaines classes de graphes. Une application importante est le problème de coupe maximum qui a pu être formulé comme un problème d'orientation. Un polytope lié à certaines matrices d'affectation qui apparaît dans les problèmes d'orientation a également été complètement décrit. Ces travaux ont donné lieu à plusieurs publications, dont [18.J.mt, 46.C.mt, 34.C.mt, 25.C.mt].

Un travail avec l'université de Caen s'est focalisé sur les problèmes de coupes étendues et coupes disjointes. Nous avons étudié le problème de calcul de coupes (dans un graphe) étendues. Nous faisons également le lien avec le problème de  $r$  coupes disjointes et déterminons le dominant de l'enveloppe convexe en généralisant ainsi le résultat classique où  $r = 1$  correspondant au dominant de l'enveloppe convexe des coupes. Plusieurs publications ont découlé de cette collaboration, dont [7.J.mt].

Un travail commun avec Orange Labs a porté sur les graphes à composantes connexes unicycliques où nous avons étudié le problème de conception de réseaux à composantes connexes unicycliques avec une contrainte sur la longueur des cycles. Des méthodes polyédrales ont été déployées pour concevoir des algorithmes à plans coupants.

Toujours dans le cadre de collaborations avec Orange Labs, nous avons aussi travaillé sur les problèmes de flot avec une nouvelle technique de génération de colonnes pour résoudre le problème de flot concurrent maximum. Lorsque toutes les demandes ont la même source, un algorithme combinatoire fortement polynomial est proposé. Dans un autre travail, nous proposons également un algorithme polynomial pour écouler un flot maximum lorsque nous routons sur des paires de chemins ne contenant pas d'arcs communs susceptibles de tomber en panne [19.J.mt].

### **Optimisation robuste**

L'équipe travaille également sur l'optimisation robuste. Une nouvelle approche intitulée optimisation robuste multi-polaire vient d'être proposée dans le cadre d'une collaboration avec des chercheurs d'Orange Labs et de l'université de Technologie de Varsovie. Cette approche généralise d'autres approches de la littérature et permet par exemple de montrer que l'approche affine est optimale lorsque l'ensemble d'incertitude est un simplexe.

### **Programmation quadratique**

Une collaboration avec Orange Labs et l'université nationale australienne, Canberra, a porté sur des problèmes d'optimisation qui font intervenir des contraintes quadratiques non-convexes. Nous avons proposé de nouvelles approches pour approcher ces fonctions quadratiques par leurs enveloppes convexes sur l'hypercube.

### **Applications réseaux**

Un travail avec l'équipe R3S a porté sur l'optimisation d'un système d'adressage, où nous avons proposé une nouvelle technique d'adressage et de résolution de noms se basant sur un arbre dont les feuilles contiendraient les informations utiles. Nous optimisons les performances de la méthode en jouant sur le nombre de fonctions de hachage qui définit le taux de réplication des informations.

Dans le cadre de travaux sur l'optimisation des réseaux optiques avec Orange Labs, une approche par programmation linéaire en nombres entiers a été proposée pour étudier plusieurs variantes de problèmes d'optimisation de réseaux FTTH (Fiber to the Home), incluant les types de câbles à installer, les opérations de splicing et de tapping à déterminer, etc. Des algorithmes d'approximation ont été également proposés dans certains cas.

Un travail commun avec CentraleSupélec a porté sur l'allocation de ressources dans les réseaux SC-FDMA (Single Carrier - Frequency Division Multiple Access). Le problème se ramène à l'affectation de canaux consécutifs à chaque utilisateur. Nous le résolvons par les approches MIP (Mixed-Integer Programming), donnons un résultat d'inapproximabilité et exhibons certains cas où le problème est soluble en temps polynomial. D'autres travaux sur les problèmes d'affectations de ressources dans les réseaux sont menés en collaboration avec des collègues de l'université de Reggio Emilie en Italie.

Nous avons également considéré un problème de partage de taxis qui a été formulé par un programme linéaire en nombres entiers. La complexité et l'approximabilité du problème ont été étudiées. Dans un autre travail, nous nous sommes intéressés au problème de cablage d'un réseau d'éoliennes [4.J.mt].



Dans une collaboration avec l’université Libre de Bruxelles, nous travaillons sur des problèmes de chimio-informatique, et spécifiquement sur des réseaux de pétri pour trouver des routes réactionnelles à moindre coûts, et pour synthétiser des molécules de la manière la moins coûteuse [5.C.mt].

### **Optimisation distribuée**

Une activité importante, en collaboration notamment avec Telecom ParisTech, porte sur l’optimisation distribuée et le consensus. Dans un réseau où chaque nœud a sa propre fonction objectif, nous avons cherché un consensus qui permet de minimiser la somme de ces fonctions. En introduisant un terme additif de variation totale, nous montrons que l’algorithme Gossip classique permet de résoudre le problème de manière distribuée sous certaines conditions générales. Il peut même résister à des agents têtus ou malveillants. Ceci a donné lieu aux publications suivantes : [12.J.mt, 47.C.mt].

Dans ce même axe, nous travaillons aussi sur l’optimisation distribuée et plus spécifiquement sur l’étude de la convergence d’algorithmes d’optimisation distribuée asynchrone basée sur le Gossip qui a donné lieu à la publication suivante : [21.J.mt], ainsi qu’une thèse Digitéo et qui a donné lieu, entre autres, à la publication suivante : [15.C.mt]. Et enfin une thèse Futur et Rupture de l’Institut Mines-Télécom sur l’analyse de la convergence d’algorithmes de consensus, ici le Gossip géométrique dans les espaces métriques CAT(k), qui a donné lieu, entre autres, aux publications suivantes : [13.J.mt, 41.C.mt].

### **Dynamique des opinions**

Une collaboration avec le groupe Performance, l’équipe R3S et l’INRIA a porté sur la dynamique d’opinion dans les réseaux sociaux, notamment : i. l’analyse de la convergence d’un algorithme de dynamique d’opinion, dans le cadre d’une thèse Futur et Rupture de l’Institut Mines-Télécom, et qui a donné lieu à la publication suivante [42.C.mt], ii. des problèmes de compétition dans un réseau social en faisant certaines hypothèses sur la propagation des opinions, et où nous avons utilisé des outils d’optimisation continue, d’optimisation combinatoire, d’optimisation robuste et de théorie des jeux, avec comme application une campagne électorale suivie d’un vote. Ceci a donné lieu à plusieurs publications, dont [3.C.mt].

### **Analyse des données**

Nous travaillons aussi sur l’analyse des données. Dans le cadre par exemple de Teralab, la plateforme de services Big Data de l’Institut Mines-Télécom, nous avons travaillé sur des algorithmes de détection de fraude qui ont donné lieu à des contrats bilatéraux avec une entreprise d’assurance et une entreprise d’acheminement des colis. Mentionnons également un travail avec Telecom ParisTech sur la détection d’anomalies [44.C.mt], et un travail avec l’équipe ACMES sur l’analyse de données de mobilité [21.C.mt].

Nous travaillons aussi sur la conception d’algorithmes de ranking et de recommandation. Les problèmes de ranking sont abordés sous l’angle de l’analyse de Fourier, notamment dans le cadre d’une thèse Futur et Rupture de l’Institut Mines-Télécom qui a donné lieu entre autres à la publication suivante : [29.C.mt]. La conception d’algorithmes de recommandation a été menée à travers une thèse avec l’entreprise Cdiscount.

## **Bilan groupe Méthodes formelles**

Le groupe Méthodes formelles est composé des enseignants-chercheurs suivants : Ana Rosa Cavalli (professeure émérite depuis novembre 2014), John Paul Gibson, Stéphane Maag, et Amel Mammar, ainsi que Natalia Kushik depuis décembre 2015, et Guillaume Burel, Catherine Dubois et Renaud Rioboo depuis janvier 2016. Les contributions principales des membres du groupe Méthodes formelles concernent les deux domaines suivants : vérification formelle et preuves (G. Burel, C. Dubois, J. P. Gibson, A. Mammar, R. Rioboo), et techniques de test et monitoring (A. Cavalli, N. Kushik, S. Maag).

### **Vérification et preuves**

Dans l’axe vérification formelle, nos contributions ont porté principalement sur les preuves automatiques pour améliorer les outils existants, en particulier en partant de la théorie de la preuve. Une partie des travaux s’inscrit dans le cadre de l’utilisation de FoCaLiZe (<http://focalize.inria.fr/>) qui permet de spécifier, implanter et prouver. Les preuves de FoCaLiZe sont faites à l’aide de Zenon (<http://zenon.inria.fr>) qui a été adapté pour raisonner modulo un système de réécriture, permettant ainsi de remplacer certaines étapes logiques par des étapes de calcul (zenon Modulo). Un nouveau backend pour FoCaLiZe vers Dedukti a été développé. Un résultat important est la réalisation via FoCaLiZe, Zenon modulo et Dedukti d’une preuve combinant un développement Coq et une preuve extraite de la bibliothèque standard de HOL. Par ailleurs, nous avons proposé une approche offrant des garanties de correction sur les produits issus d’une ligne de produits logicielle. La méthode s’appuie sur un langage, FFML, inspiré de FoCaLiZe. Le composeur développé permet de générer automatiquement des produits corrects

par construction à partir d'une configuration valide. Ces travaux ont donné lieu notamment aux publications suivantes [12.C.mt, 22.C.mt, 20.C.mt, 1.J.mt].

Nous avons aussi étudié les liens entre démonstration automatique et théorie de la preuve. D'une part, nous avons étudié comment obtenir de meilleures méthodes de démonstration automatique à travers les cadres de la déduction modulo théorie et de la focalisation. Ceci nous a conduit à continuer le développement du prouveur *iProverModulo*. D'autre part, nous nous sommes intéressés à la généralisation de preuves à partir de prouveurs automatiques, en particulier en produisant des preuves pour le vérificateur de preuve universel *Dedukti*. *iProverModulo* est donc maintenant capable de produire de telles preuves. Ces travaux se font en collaboration avec l'équipe *Deducteam* du LSV/Inria.

Un autre axe concerne l'application de la vérification formelle. Dans ce contexte, nous avons contribué principalement à la prise en compte des règles de contrôle d'accès dans le domaine des systèmes d'information [11.J.mt, 13.C.mt, 16.J.mt]. Dans le cadre de la mise en place du dossier médical informatisé en collaboration avec l'université de Sherbrooke, un langage formel a été proposé pour l'expression des règles d'accès aux données du patient ainsi qu'un ensemble d'algorithmes prouvés pour la gestion du contrôle d'accès aux dossiers. L'approche a été validée par un outil. Par ailleurs, nous avons également travaillé sur la prise en compte des règles dynamiques de contrôle d'accès pour les applications système d'information en proposant une approche MDE (*Model-Driven Engineering*) qui permet de considérer de telles règles dès les premières phases de développement d'un système. En collaboration avec l'équipe *ACMES*, une approche de modélisation et vérification de l'allocation de ressources et de la configuration de processus dans le Cloud a été proposée [8.J.mt]. En collaboration avec l'équipe *R3S*, nous avons abordé la sûreté des véhicules autonomes. D'autres travaux ont porté sur l'application de la vérification formelle au business process management. Ils ont été publiés dans les conférences suivantes [24.C.mt, 23.C.mt, 35.C.mt, 36.C.mt], dont le prix du meilleur papier à la conférence internationale *RCIS* 2016.

En utilisant notre expérience dans le domaine de la vérification et de la validation de logiciels et des services, nous avons contribué au développement basé sur les modèles. Dans ce contexte, nous avons étudié le vote électronique et nous avons mis en avant une proposition de recherche sur une ligne de produits logiciels pour les systèmes de vote électronique qui combinent de plus de façon complémentaire théorie, pratique et pédagogie [10.C.mt].

### **Techniques de test et monitoring**

Dans cette thématique, une première partie des travaux a porté sur la proposition de nouvelles méthodologies de génération de tests à partir de modèles formels ou l'amélioration de méthodologies existantes, ainsi que l'étude de leur complexité. Les principaux résultats obtenus pendant cette période sont : i) l'identification des machines à états finis non déterministes ayant une complexité réduite des problèmes d'identification d'état et donc, une complexité réduite pour la génération de tests, et ii) des nouveaux modèles de faute et des techniques de test pour les infrastructures programmables [5.J.mt]. Une collaboration avec l'équipe *R3S* a pour cadre une thèse *Futur et Rupture* de l'Institut Mines-Télécom qui a porté sur l'analyse et la validation des requêtes réseau virtuelles [8.C.mt].

Une deuxième partie des travaux a porté sur le monitoring en ligne. Nous avons travaillé dans cet axe sur la modélisation des protocoles de communications à partir d'automates symboliques. L'objectif principal visait l'exécution automatisée de suites de test. Or, dans de nombreux cas d'étude réels, les données et les paquets échangés dans des réseaux de communication complexes sont très importants. Nous avons mis en place des techniques pour éviter les explosions combinatoires [39.C.mt, 20.J.mt]. Nous travaillons aussi depuis de nombreuses années avec des collègues de l'université d'Orsay et notamment le laboratoire *LRI* et l'équipe *VALS*. Les travaux se concentrent sur les approches de monitoring fonctionnel et le test passif des protocoles de communication. La formalisation des propriétés à superviser et la prise en compte de la causalité ont été mis en oeuvre [26.C.mt, 14.C.mt]. Nous collaborons aussi avec deux laboratoires étrangers, le laboratoire *RLICT* de l'université Galileo à Guatemala City et le Laboratoire du Département d'Informatique de l'université *JiaoTong* de Pékin. Les travaux portent sur la validation des protocoles sans fil utilisés dans les *MANETs* (*Mobile Ad Hoc Networks*) mais aussi dans l'*IoT* (*Internet of Things*). Nous étudions la validation (conformité et interopérabilité) des protocoles, notamment de routage [37.C.mt, 9.J.mt, 20.J.mt].

Enfin, une troisième partie des travaux a porté sur la qualité et la sécurité du logiciel et des services. Nous avons développé une activité de recherche autour du test de services web et leur composition et de l'analyse de la *QoS* (*Quality of Service*) des services afin d'évaluer la *QoE* (*Quality of Experience*) des usagers. Aussi, nous avons proposé des techniques pour l'estimation de la qualité de service basées sur les machines à états finis étendues. Nous avons développé de nouvelles heuristiques pour l'optimisation du code source. Ceci a donné lieu notamment aux publications suivantes : [2.C.mt, 7.C.mt, 6.C.mt, 30.C.mt, 43.C.mt].

## Groupe Performance

Le groupe Performance est composé des enseignants-chercheurs suivants : Tulin Atmaca, Monique Becker (professeure émérite depuis novembre 2013), Hind Castel et Tijani Chahed. Le bilan scientifique du groupe s'articule autour des problématiques liées à l'étude de performance dans les réseaux cellulaires (4G, 5G), les réseaux optiques et le Cloud notamment.

### Allocation de ressources dans les réseaux mobiles

Une activité importante concerne l'allocation de ressources et le partage de charge dans le cadre des réseaux cellulaires (4G et 5G notamment). Des études de modélisation et d'évaluation des performances sont menées afin de prendre en compte l'équilibrage de charge et les techniques de remédiation aux interférences telles que EICIC (Enhanced Inter Cell Interference Cancellation) ou COMP (Coordinated Multipoint). Une collaboration récente avec l'université Paris-Sud et l'université Paris Descartes porte sur l'allocation de ressources dans le C-RAN (Cloud-Radio Access Network) partagés entre plusieurs opérateurs mobiles virtuels (MVNO) en utilisant les enchères [2.J.mt].

Dans le contexte des réseaux HetNet (Heterogeneous Networks), composés de macros et petites cellules, une collaboration avec l'université Libanaise et Orange Labs a porté sur l'étude du multi-homing, avec la possibilité pour un flux d'utiliser de manière simultanée un accès WiFi et un accès cellulaire [11.C.mt]. Une autre collaboration avec Orange Labs dans le cadre d'une thèse Cifre a porté sur le déploiement de petites cellules, statiques ou mobiles, voir même en drones, pour écouler le trafic provenant de hotspot, avec une modélisation mixant les niveaux flots et paquets [10.J.mt, 17.C.mt]. Enfin, une troisième collaboration avec l'université Libanaise et Orange Labs a porté sur l'utilisation jointe des réseaux LTE (Long Term Evolution) et des réseaux DVB (Digital Video Broadcast) pour la transmission efficace de la télévision numérique mobile. La valeur de Shapley, un concept issu de la théorie des jeux coopératifs, a été utilisée pour le partage des revenus et des coûts entre ces deux types d'opérateurs [22.J.mt].

Nous nous sommes aussi intéressés à la partie métré d'un réseau optique pour montrer dans quelle mesure cette architecture peut répondre aux exigences des prochaines technologies comme la 5G. Une évaluation des performances a été réalisée et il a été proposé une nouvelle alternative de gestion et de transmission des paquets pour la commutation de paquets optiques en couleur. Cette alternative pourrait être utilisée pour différents modes de transmissions (broadcast&select et unicast). Contrairement aux autres méthodes classiques, l'approche exploite une caractéristique spécifique d'architecture de nœud qui est le surdimensionnement des ressources. La méthode proposée a été évaluée sous différentes conditions et contextes (anneau unidirectionnel et bidirectionnel). Nous avons évalué l'efficacité de cette méthode en utilisant le modèle de trafic Poissonien ainsi qu'un trafic plus réel tels que les traces de trafic provenant de CAIDA (Anonymized Internet Traces Dataset), un modèle caractérisé par un trafic bursty. Les résultats semblent intéressants dans un certain nombre d'autres applications nécessitant des taux de transmission extrêmes de données qui ont un profil de trafics divers (par exemple en rafale), comme on peut en trouver dans les grands centres de données ou les centres commerciaux. Ces travaux ont donné lieu à plusieurs publications, dont : [16.C.mt].

Dans le cadre d'un contrat industriel bilatéral avec Orange Labs, nous avons travaillé sur la notion d'impatience dans les réseaux mobiles, 4G notamment, et son impact sur la performance des réseaux et des usagers. Ceci a donné lieu à la publication suivante : [28.C.mt].

Une collaboration avec Sup'com, Tunis, a porté sur la notion d'agrégation de la capacité dans le réseau backhaul en utilisant les matching games, un outil issu de la théorie de jeux coopératifs, qui permet de trouver la meilleure correspondance entre les points d'accès avec des ressources en trop et ceux qui sont en manque, avec plusieurs variantes de matching : one-to-one, one-to-many et many-to-many [18.C.mt, 31.C.mt].

### Economie d'énergie

Plusieurs travaux ont porté sur la notion d'énergie dans les réseaux mobiles. Un premier travail, dans le cadre d'une thèse Cifre avec Orange Labs, a porté sur la modélisation de l'énergie par catégorie de service en utilisant ici aussi la valeur de Shapley. Un deuxième travail a porté sur l'interaction entre un opérateur de télécommunications et un smart grid en utilisant une modélisation basée sur les MDP (Markov Decision Processes) afin de trouver la politique optimale pour l'opérateur afin de réduire au maximum sa facture énergétique, sur le court et long termes. Un troisième travail a étudié la notion de sleep mode dans les réseaux hétérogènes (HetNet), composés de petites cellules en plus de la station de base macro. Ici aussi nous avons dérivé la politique optimale pour réduire la consommation globale d'énergie en utilisant les MDP, avec information complète, partielle ou encore retardée [3.J.mt, 9.C.mt, 1.C.mt, 32.C.mt, 19.C.mt].

Nous nous sommes aussi intéressés à l'analyse de la performance et la consommation énergétique dans le Cloud.

Nous avons utilisé pour cela des modèles de files d'attente avec seuils, de type hystérésis. Ce type de modèle de file d'attente permet de représenter la dynamique des ressources en fonction de la charge. Le système est représenté par une Chaîne de Markov dont la complexité augmente assez vite avec le nombre de serveurs (ou ressources). Différentes méthodes analytiques et numériques ont été utilisées et elles ont été comparées en termes de rapidité d'exécution et de précision. Afin d'analyser le compromis entre performance et énergie, une mesure globale est définie prenant en compte à la fois les performances et la consommation énergétique, et des heuristiques sont développées afin de déduire les valeurs de seuils correspondant à la mesure globale optimale.

### **Propagation de l'information**

Nous avons mené des travaux sur les réseaux DTN (Delay Tolerant Networks) en collaboration avec l'université de Nice et l'INRIA. Nous avons proposé un nouvel algorithme pour le transport fiable des messages, en utilisant le codage réseau, ainsi que le tracking et l'estimation de degré de propagation des messages [17.J.mt].

Dans le cadre d'une thèse Futur et Rupture de l'Institut Mines-Télécom, un travail en collaboration avec l'INRIA et l'université de l'Illinois à Urbana-Champaign a porté sur la modélisation de la propagation de contenus, de manière compétitive, entre plusieurs usagers dans les réseaux sociaux en utilisant les processus de Hawkes [?].

### **Problèmes de files d'attente**

Une étude que nous avons mené concerne l'analyse de files d'attente avec des lois d'arrivée et de service générales, décrites par des histogrammes de traces de trafic réelles. Un trafic à phase a été supposé, ce qui rend l'analyse de la file d'attente très complexe à cause de l'explosion du nombre d'états. Des techniques d'agrégations bornantes des histogrammes ont permis de résoudre le problème de calcul des mesures de performance, et des compromis ont été proposés entre précision des résultats et complexité de calculs. Ces travaux ont donné lieu à plusieurs publications, dont [48.C.mt, 15.J.mt, 27.C.mt].

Enfin, une collaboration avec l'université de Californie à Santa Cruz s'est intéressée à l'analyse de files d'attente multi-serveurs avec politique à seuils, lois d'arrivée et de service générales. Ce modèle permet de représenter un système avec différents niveaux de fonctionnement, où chaque niveau représente un certain nombre de serveurs actifs. L'analyse du système se fait par une approche mathématique, basée sur la définition des équations de balance et une approche semi-numérique est proposée pour obtenir la distribution de probabilités stationnaires. Des études de systèmes tels que des Clouds permettent d'obtenir des résultats numériques sur des mesures de performance et de consommation énergétique, permettant de proposer un compromis entre les deux [6.J.mt, 14.J.mt].

## **5.2.2 Faits marquants de l'équipe METHODES**

Nous relevons, pour la période 2013-2018, les faits marquants suivants :

- En 2016 nous avons accueilli 7 nouveaux membres de l'ENSIIE : 4 permanents et 3 associés.
- Jérémie Jakubowicz a occupé l'unique chaire CNRS du laboratoire SAMOVAR de 2011 à 2016. Il a assumé, de 2013 à 2017, à temps partiel (20%), la fonction de responsable de la data science au sein de la plateforme dédiée aux services Big Data de l'Institut Mines-Télécom : Teralab.
- Les publications en revues de l'équipe représente 30% du nombre de publications globale, dont 42% classés A\*/A selon le classement Era Core.
- Nous avons reçu les prix de meilleurs papiers aux conférences suivantes : ENASE 2013, HPCC 2013, RCIS 2016, et CN 2018.
- 34 thèses ont été soutenues.

## **5.3 Organisation et vie de l'équipe METHODES**

La vie de l'équipe s'organise autour de réunions régulières et de séminaires individuels ou lors de journées organisées par les membres de l'équipe ainsi que des membres invités.

Pour ce qui est de l'utilisation du budget de l'équipe, nous avons privilégié de financer les missions des membres de l'équipe, permanents et doctorants, pour des conférences de rang A en majorité.

## **5.4 Analyse SWOT de l'équipe METHODES**

L'analyse de nos activités, de nos résultats et de notre positionnement conduit à l'auto-évaluation suivante :

Points forts	<ul style="list-style-type: none"> <li>— L'équipe METHODES occupe un rôle central au sein du laboratoire Samovar grâce à l'expertise de ses membres en méthodologies et outils mathématiques et de simulations.</li> <li>— L'équipe a bénéficié de l'arrivée de huit nouveaux membres.</li> <li>— Les publications de l'équipe METHODES ont continué à apparaître dans des revues et des conférences internationales de premier plan.</li> <li>— L'équipe continue à assurer un bon niveau de financement, avec des sources diversifiées, tels que les bourses Cifre, les projets ANR et PIA, les projets Européens, les projets bilatéraux avec les industriels, et les financements Institut Mines-Télécom.</li> <li>— Les membres de l'équipe METHODES sont très actifs dans les rôles d'animation scientifique dans les comités éditoriaux de revues, l'organisation de conférences, et différents groupes de travail, ainsi que dans les rôles d'évaluation et de conseil.</li> <li>— Les membres de l'équipe participent à différentes formations doctorales et masters de recherche.</li> </ul>
Points faibles	<ul style="list-style-type: none"> <li>— L'équipe, comme le laboratoire, n'a pas encore réussi à recruter de nouveaux membres permanents CNRS.</li> <li>— L'équipe manque de moyens propres qui lui donneraient la possibilité d'initier et de renforcer certaines thématiques et actions de recherche.</li> </ul>
Opportunités	<ul style="list-style-type: none"> <li>— Le déménagement partiel prochain de certains membres du laboratoire et l'intégration dans la nouvelle structure NewUni, constituent des nouvelles opportunités de rapprochement et de collaboration avec les différents acteurs, de premier plan, qui seront présents sur le site.</li> <li>— Le positionnement thématique de l'équipe, centré sur les méthodes, permettra une interaction plus facile avec les différents partenaires travaillant aussi bien sur les méthodes que sur leurs applications dans différents contextes.</li> </ul>
Risques	<ul style="list-style-type: none"> <li>— La bi-localisation prochaine du laboratoire et de l'équipe, dans le cadre du déménagement de quelques uns de ses membres à Saclay indiqué ci-dessus, pourrait créer quelques difficultés pour la vie de l'équipe, tels que l'organisation de réunions, séminaires, ainsi que l'interaction entre ses membres.</li> </ul>

## 5.5 Projet scientifique de l'équipe METHODES

### 5.5.1 Projet groupe Optimisation

Nous poursuivrons nos travaux en optimisation combinatoire et en programmation mathématique. Nous souhaiterions essayer de généraliser l'algorithme combinatoire qui a été proposé par l'équipe pour le problème de flot concurrent maximal dans le cas mono-source au cas multi-sources. Un autre sujet de recherche qui devrait nous préoccuper serait la description du dominant de l'enveloppe convexe de  $k$ -indépendants disjoints d'un cluster. Il s'agira là aussi de généraliser le résultat de l'équipe portant sur le dominant des  $r$ -coupes disjointes. Nous avons également démarré plusieurs investigations autour de la conjecture De Langberg-Medard proposant une condition suffisante pour pouvoir router dans un graphe non-orienté qui porte sur l'existence d'une orientation permettant d'avoir des chemins disjoints de chaque source vers toutes les destinations.

Nous étudierons aussi des problèmes de partitionnement (de type maxcut et différentes généralisations) et de couverture (problèmes liés à différents concepts de domination) dans les graphes. Il s'agira d'approfondir des approches linéaires et semi-définies pour des problèmes réputés difficiles et d'un intérêt majeur pour la communauté scientifique en optimisation combinatoire.

Nous travaillerons également sur des problèmes de dimensionnement dans les réseaux (recherche de routes optimales de véhicules partagés), des problèmes de chimio-informatique et d'optimisation dans les réseaux électriques.

Enfin, nous poursuivrons l'étude de l'approche multi-polaire pour l'optimisation robuste proposée récemment par l'équipe et nous évaluerons son apport dans le cadre d'applications liées aux réseaux.

## 5.5.2 Projet groupe Méthodes formelles

Les membres de l'équipe continueront à travailler dans les deux axes : vérification et preuves, et test et monitoring, ainsi que leur application pour le génie logiciel formel.

Dans l'axe vérification et preuves, nous nous concentrerons surtout sur les applications des outils de vérification, y compris en temps réel. La validation et le test se trouveront notamment dans la vérification des propriétés de sécurité et de détection d'intrusion, la certification des systèmes complexes, et les lignes de produits et vote électronique.

Nous travaillerons sur deux axes liés à la modélisation et le développement formels d'outils de détection d'intrusion dans les systèmes bancaires mais également sur la prise en compte d'ontologies pour l'amélioration du processus de modélisation des systèmes complexes et critiques. Pour la modélisation des outils de détection d'intrusions, l'utilisation d'une représentation graphique de l'attaque est préconisée, comme les diagrammes états/transitions, à partir de laquelle un programme prouvé est généré formellement pour sa reconnaissance. Cette approche sera basée sur l'utilisation de la méthode formelle B/EventB. Nous explorerons également l'utilisation des ontologies pour la conception de systèmes complexes et critiques afin de décrire la connaissance du domaine en termes de concepts et de relations. L'intérêt majeur d'utiliser une ontologie est de développer un modèle pivot partagé par tous les acteurs d'un domaine donné favorisant ainsi la réutilisation.

Toujours dans cet axe, nous continuerons à travailler sur le développement et les applications de FoCaLiZe. Une application importante concerne la certification formelle d'une famille de validateurs XML. Après une première étude, est apparue la nécessité d'étendre le compilateur de FoCaLiZe de manière à gérer plus aisément la récursion mutuelle. Ensuite, l'interopérabilité des preuves et des prouveurs : après la première preuve de concept, il s'agit de généraliser l'approche et de mesurer son passage à l'échelle. Ce travail se poursuivra en collaboration avec l'équipe-projet INRIA Deducteam.

Nous poursuivrons aussi l'exploration des liens entre démonstration automatique et théorie de la preuve. Il s'agit de faire coopérer étroitement le vérificateur de preuve Dedukti avec des outils de démonstration automatique, de façon à alléger le travail de construction interactive de la preuve. L'objectif est de pouvoir utiliser Dedukti comme liant pour l'interopérabilité entre différents outils de démonstration, qu'elle soit automatique ou interactive. Il faut pour cela également augmenter l'expressivité de Dedukti pour lui permettre de prendre en compte plus de théories, notamment grâce à la polarisation.

En ce qui concerne le génie logiciel formel, nous envisageons de travailler sur les axes suivants : le génie logiciel formel et notamment l'interaction des services, le vote électronique, les méthodes formelles et l'apprentissage par problèmes, et enfin, les lignes de produits logiciels. Nous bénéficierons ici des techniques de vérification et validation pour le développement basé sur les modèles, en relation avec le premier axe de ce groupe de travail.

Dans l'axe test et monitoring, nous nous intéresserons aux deux activités : monitoring et test. Pour les techniques de monitoring, notre projet s'articule autour de deux axes principaux. Le premier axe concerne le test des systèmes communicants et principalement l'Internet des objets (IoT) en prolongement de nos travaux sur les MANET ainsi que les véhicules autonomes (travail commun avec l'équipe R3S). Le monitoring distribué fonctionnel, l'interopérabilité des protocoles et le test actif en utilisant TDL sont les trois pistes majeures dans ce domaine. Ce langage pourrait par ailleurs permettre de décrire des règles utilisées dans le test passif. Les outils de virtualisation sont en outre des solutions envisagées pour éprouver les méthodes qui seront définies. Le deuxième axe se réfère à la mesure logicielle en y intégrant des techniques d'apprentissage non-supervisé. Nous envisageons également de faire converger nos études sur le monitoring avec nos travaux sur la mesure logicielle et de transférer les résultats dans l'industrie (suite notamment aux collaborations déjà tissées).

Pour les techniques de test, nous prévoyons de continuer le travail sur le développement des nouvelles techniques de test basées sur les modèles formels ainsi que l'amélioration des méthodes existantes de la certification des différents systèmes complexes, y compris les infrastructures programmables (travail commun avec l'équipe R3S). Dans ce cas, il est prévu que les techniques d'apprentissage soient utilisées d'une part, pour la génération des séquences de test et, d'autre part, pour la vérification en ligne des composants de réseaux. Les problèmes de complexité pour les techniques de test et de vérification basés sur les modèles formels seront également étudiés ; en conséquence, des nouvelles approches pour réduire cette complexité seront proposées (il est également possible que l'Intelligence Artificielle puisse être utile pour de nouvelles heuristiques). Les modèles de faute originaux ainsi que les représentations évolutives pour les spécifications formelles permettront de fournir des nouvelles heuristiques pour le test.

### 5.5.3 Projet groupe Performance

Le projet du groupe Performance s'inscrit dans la lignée des thématiques indiquées ci-dessus, à savoir, l'évaluation de performance, l'allocation de ressources et l'étude du compromis entre performance et consommation d'énergie dans les réseaux mobiles, notamment 5G, le Cloud et les réseaux optiques (notamment d'accès).

Dans les réseaux 5G, nous nous intéresserons à la notion de slicing pour le transport de différentes catégories de service, notamment les services dits critiques, pour un nombre massifs d'objets connectés, appelé MTC (Machine Type Communications), avec ses deux variantes URLLC (Ultra Reliable Low Latency Communications) et mMTC (massive MTC).

Nous continuerons aussi nos travaux sur l'optimisation de l'énergie dans ce type de réseau, en utilisant des notions d'apprentissage et d'intelligence artificielle qui permettront d'apprendre les paramètres du système ainsi que les comportements des objets connectés. La notion d'économie d'énergie est aussi envisagée sous la forme des sleep modes avancés qui définissent dans le cadre de la 5G des durées très petites de sleep visant des petites composantes de la station de base notamment.

Dans le cadre des réseaux Clouds, il est nécessaire d'analyser des politiques de gestion efficace des ressources dans le Cloud afin de garantir à la fois les performances et l'économie d'énergie. Nous étudierons en particulier des mécanismes basés sur les modes d'activation/désactivation des ressources, les migrations de machines virtuelles, l'équilibrage de charge, les consolidation de serveurs physiques, etc.

L'évaluation des politiques et la recherche de politiques optimales pour le fonctionnement des systèmes tout en économisant l'énergie se fera par des modèles mathématiques tels que les Chaînes de Markov, les MRP (Markov Reward Process), et les MDP (Markov Decision Process). Comme ces deux grandeurs évoluent en sens inverse (quand on améliore les performances, on dégrade typiquement la consommation énergétique, et réciproquement), la recherche de la politique ou des paramètres permettant d'obtenir la valeur optimale est complexe. Une méthode simple serait de procéder par une recherche exhaustive, mais cela prendrait trop de temps même si on fait appel à des algorithmes de résolution numérique efficaces. D'autres méthodes proposent des heuristiques intéressantes, mais dans ce cas il est parfois difficile d'évaluer la précision des résultats. C'est pourquoi il peut être intéressant de faire appel à des techniques d'agrégation de l'espace d'état couplés par exemple à des MRP (qui permettent d'affecter des coûts aux états de l'espace d'état) ou à des MDP afin de calculer la mesure globale optimale d'une manière plus efficace. Ici aussi, des notions d'apprentissage seront utilisées pour apprendre les paramètres de l'environnement dans lequel évolue notre système.

# 6 Équipe R3S

## 6.1 Présentation de l'équipe R3S

### 6.1.1 Introduction

Depuis sa création en 2011, l'équipe R3S (Réseaux, Systèmes, Services, Sécurité) a évolué dans sa structuration pour aboutir aujourd'hui à deux groupes thématiques. Au cours du dernier quinquennat, elle a connu le départ du groupe technologies avancées pour les communications du futur (TAC) composé de 7 enseignants-chercheurs, la mobilité du Prof. Michel Marot de l'équipe METHODE vers R3S en 2017 et l'intégration du Dr. HDR Zonghua Zhang (IMT Lille Douai) et du Dr. Grégory Blanc en tant que membres permanents de R3S en 2016 et 2017 respectivement. Depuis 2012, l'équipe est bilocalisée entre Evry et Paris-Saclay, puisque trois de ses membres occupent actuellement des locaux dans le bâtiment Nano-INNOV.

Actuellement, l'équipe est organisée en deux groupes thématiques, à savoir :

- réseaux et services (RS) qui comprend douze membres dont huit HDR : Les travaux couvrent un large spectre depuis les protocoles de communication jusqu'aux services, en s'intéressant à la virtualisation des services et fonctions, aux réseaux contraints auto-organisés et/ou hétérogènes, aux nouvelles technologies sans fil et mobiles, au traitement de données massives, sous l'angle des performances, de la consommation d'énergie ou de l'expérience utilisateur.
- sécurité des systèmes et des réseaux (SSR) qui comprend six membres dont cinq HDR : Les travaux portent essentiellement sur la cybersécurité et la protection des données personnelles, appliquées à plusieurs cas d'usage ou applications comme les environnements virtualisés, les réseaux (SDN), les fonctions associées aux réseaux (NFV) ou l'Internet des objets (ex : SCADA).

Ces deux groupes sont impliqués dans les quatre domaines d'intérêt majeur pour l'IMT que sont l'industrie 4.0, l'énergie, les smart cities et les transports. Ils collaborent à des projets et travaux scientifiques communs, comme l'attestent les publications [43.C.rs, 54.C.rs, 33.J.rs, 32.J.rs, 35.J.rs, 19.J.rs, 28.J.rs, 29.J.rs, 40.C.rs, 52.C.rs, 67.C.rs].

Pour une meilleure lisibilité du rapport, le bilan de R3S comprend les travaux de l'ensemble des membres présents dans l'équipe. Les travaux du groupe TAC, sur toute la période, sont pris en compte dans le rapport de l'équipe TIPIC. §

### 6.1.2 Effectifs et moyens de l'équipe

L'équipe au 1<sup>er</sup> janvier 2018 est constituée de 18 membres permanents (8 professeurs, 9 maîtres de conférences, 1 ingénieur d'études), 36 doctorants, 2 post-doctorants, 18 ingénieurs de recherche (CDD), 1 ingénieur plate-forme et 2 chercheurs associés :

**Professeurs (HDR)** : Hossam Afifi, Hakima Chaouchi, Noël Crespi, Hervé Debar, Joaquin Garcia-Alfaro, Maryline Laurent, Michel Marot, Djamal Zeghlache

**Maîtres de conférences** et ingénieur d'études : Grégory Blanc, Vincent Gauthier, Marc Giroto-Genet, Badi Jouaber, Anis Laouti (HDR), Abdallah M'hamed (HDR), Hang Nguyen, Daniel Ranc, Eric Renault (HDR), Zonghua Zhang (HDR)

**Chercheurs associés** : Hassine Moun gla (HDR, Paris Descartes, depuis le 01/01/2018), Lina Nachabe Ismail (Université du Liban, depuis le 01/01/2017).



Comme le montre le tableau de l'évolution des effectifs, l'équipe est composée exclusivement d'enseignants chercheurs (aucun chercheur temps plein) et le nombre de permanents suit les fluctuations liées à la restructuration de l'équipe en 2015 et à l'intégration de nouveaux membres, telles que présentées à la section 6.1.1.

#### Evolution des effectifs de l'équipe R3S

	1/1/2013	1/1/2014	1/1/2015	1/1/2016	1/1/2017	1/1/2018
Enseignants chercheurs permanents*	22	22	15	16	18	18
Chercheurs associés*					1	2
Doctorants*	47	49	50	42	39	36
Chercheurs ou ingénieurs (CDD)**	5	7,5	12	15	21,5	10
Post-doctorants**	6	5,5	7,5	3	1,5	1

\* Effectifs présents au 1<sup>er</sup> janvier de l'année considérée

\*\* Effectifs exprimés en équivalent temps plein avec l'année 2018 ne couvrant que le premier semestre.

Nous disposons de ressources propres conséquentes avec une moyenne de 2.6 millions d'euros/an, obtenues sur des appels à projet européens et nationaux (cf. tableau des ressources de R3S). Cependant, le niveau de compétitivité de ces appels, avec des taux de réussite de moins de 15%, rend de plus en plus difficile l'obtention de financements. Ainsi, on observe un fléchissement des ressources obtenues sur les appels européens et nationaux (hors PIA), et un effort de diversification, avec l'augmentation des ressources obtenues sur des PIA et sur du mécénat grâce à notre participation à deux chaires de l'IMT. Par ailleurs, ces ressources propres sont complétées par des soutiens indirects relativement stables dans le temps, à hauteur d'une trentaine de bourses de thèses annuelles, soit un équivalent de 840 K€/an (essentiellement des CIFRE, et des bourses de gouvernement étranger).

#### Evolution des ressources en K€

ressources indirectes comprises (bourses de thèses)

		2013	2014	2015	2016	2017	Total
Contrats Rech. publics	Projets Européens	1793	1930	1710	1046	1006	7485
	Projets ANR	185	52	108	40	87	472
	PIA	38	73	108	357	200	776
	Financements publics (appels à proj.)	509	365	510	427	199	2010
	Collectivités territoriales	38	64	213	0	74	389
Contrats Rech. privés	Contrats industriels	249	278	342	181	295	1345
Financements ext. doct.	Bourses de thèse (CIFRE...) *	945	1015	980	1050	1050	5040
Institut Télécom	Fondation, mécénat	20	15	0	54	169	258
	Institut Carnot	90	35	26	0	51	202
CNRS	Soutien de base	6	11	15	11	7	50
Total		3873	3838	4012	3166	3138	18027

\* La base choisie est de 35K€/an pour une bourse de thèse.

Dans le tableau des effectifs, nous observons que le volume de doctorants suit les variations des ressources propres avec un début d'infléchissement en 2016. Par contre, cette remarque ne s'applique pas aux post-doctorants et chercheurs/ingénieurs (CDD) dont le nombre cumulé tend à croître. Cela s'explique par la difficulté d'un engagement financier sur 36 mois consécutifs. Par ailleurs, il est utile de clarifier la notion de chercheurs/ingénieurs (CDD). Il peut s'agir d'ingénieurs de développement, ou bien de chercheurs qui effectuent un vrai travail de recherche, mais qui n'ont pas pu prendre le statut de post-doctorant pour des raisons administratives. La proportion de chercheurs est de 45%, ce qui représente une force de recherche d'environ 7 temps plein par an. Les doctorants et post-doctorants manifestent souvent l'envie de prolonger leur expérience de recherche avec les chercheurs de l'équipe. Ainsi, 8 doctorants sont restés en post-doctorat pour une durée moyenne de 13 mois et 9 post-doctorants ont prolongé leur séjour en tant que chercheurs (CDD) pour une durée moyenne de 14 mois.

L'équipe est très engagée dans la formation à la recherche. Elle a participé à former 106 doctorants pour une durée de thèse de 41,09 mois en moyenne. Elle a accueilli 26 post-doctorants et 79 chercheurs/ingénieurs (CDD) pour une durée moyenne de 11 mois.

**Evolution des modes de financement des thèses de doctorat**  
(exprimés en nombre de bourses de thèses)

	2013	2014	2015	2016	2017	2018
CIFRE	11	13	15	15	12	11
CEA	1				1	2
Gouvernement étranger	3	6	7	9	8	6
Gouvernement français	2	2	1	1	1	1
Bourse UPMC		1	1	1		
Contrats européens	14,5	14,5	16	11	10	6
Projets Pôle de comp. Systematic	2	3	2	2	2	2
Contrat ANR	1					
Institut Telecom	2,5	2,5	4	2	3	4
Salarié de l'IRT SystemX	1	1	0	1	1	1
Autofinancement (salariés)	9	6	4	3	1	3

### 6.1.3 Profils d'activité

L'équipe est impliquée dans des activités de recherche, de valorisation, d'animation et de réflexion sur la société, en partenariat avec des industriels, des académiques et des institutions, et ce à différentes échelles internationales, européennes, nationales et régionales.

Même s'il est difficile de cloisonner les activités entre elles, nous avons réalisé l'exercice et abouti à la synthèse suivante :

- activité scientifique (50% de l'activité environ) : cette activité concentre les efforts de l'équipe avec une production scientifique qui prend des formes variées et qui peut être synthétisée par les chiffres suivants : 167 publications dans des journaux et 341 en conférences (cf. section 6.2), 7 prix du meilleur article scientifique, édition de 10 ouvrages scientifiques, 4 logiciels produits et disponibles sur un github, 3 bases de données publiés, 5 plates-formes, 4 prototypes et 4 séjours invités dans des laboratoires étrangers.
- activité de valorisation et de transfert (20% environ) : cette activité est soutenue au travers d'une quinzaine de projets européens (dont 5 projets FP7, 4 projets H2020), 31 CIFRE, 3 projets FUI, 9 contrats industriels, 2 chaires de l'IMT et 2 PIA, ce qui nous permet d'entretenir des relations privilégiées avec de grandes entreprises et des PME/TPE et de bénéficier d'un chiffre d'affaire annuel moyen de 2.6 millions. Nous avons publié 4 brevets (acceptés) et avons contribué à la standardisation dans les organisations ETSI, oneM2M et ITU-T. Par ailleurs, nous nous sommes beaucoup investis auprès des instances publiques en prenant part au PPP ECSO (Public-Private Partnership European Cyber Security Organisation) et à plusieurs consultations nationales sur le numérique (Ministère de l'Economie et des finances, Conseil National du Numérique).
- formation par la recherche (15% environ) : l'équipe a formé 106 doctorants, 62 d'entre eux ayant soutenu leur thèse de doctorat, ce qui revient à une moyenne sur la période de 5,5 ans, d'environ 9 doctorants par HDR (resp. 6 doctorants par permanent), soit environ 4 doctorants simultanés par HDR (resp. 2,5 doctorants par permanent). De plus, nous avons encadré 26 post-doctorants et 85 chercheurs/ingénieurs (CDD), en plus d'étudiants en stage de M1 et M2. Nous avons aussi contribué activement aux formations de masters de l'Université Paris-Saclay et de Télécom Sudparis, ce qui peut être synthétisé comme suit : montage de 3 masters de l'Université Paris-Saclay, coordination du M1 Computer Science (International Track, mention informatique) et coordination d'une dizaine de modules de cours de masters (M1 et M2).
- appui à la communauté (15% environ) : cette activité couvre l'organisation de plus d'une dizaine de conférences et workshops, dont la conférence de premier plan en sécurité RAID 2016, notre participation à des comités éditoriaux (Computers & Security, Computers & Electrical Engineering, ...), des comités de pilotage de conférences, l'animation de journées scientifiques thématiques soutenues par l'IMT ou un GDR (Pré-GDR Sécurité informatique, GdR RSD). Nous sommes régulièrement sollicités pour des commissions d'évaluation de projets H2020 ou FUI en tant qu'expert, et aussi pour des évaluations de projets ANR ou de dossiers de chercheurs.

## 6.2 Produits de la recherche et activités de la recherche de l'équipe R3S

Le tableau suivant présente le nombre de publications scientifiques sur la période 2013-2018, en distinguant les publications classées selon le référentiel ERA-CORE (rangs A et A\* confondus et B) et le référentiel ISI. Ces référentiels approuvés par le conseil de laboratoire de SAMOVAR ont servi de guide de bonne pratique de publication auprès des membres de l'équipe pendant toute la période. Notons qu'une publication n'est comptabilisée qu'une seule fois dans le tableau. Seules les publications des membres de R3S présents après la restructuration de l'équipe en 2015 sont prises en compte. Pour les chercheurs associés, seules les publications copubliées avec un membre permanent sont considérées.

**Volume de publications de R3S**

	2013	2014	2015	2016	2017	2018	Total
Rev A* et A	4	7	7	5	5	5	33
Rev B	11	8	5	3	6	1	34
Rev ISI	6	6	9	12	14	4	51
Total Rev classées	21	21	21	20	25	10	118
Total Rev	24	21	31	32	42	12	162
Conf A* et A	6	4	4	7	11	1	33
Conf B	9	16	24	24	20	3	96
Conf ISI	23	12	14	18	16	6	89
Total Conf classés	38	32	42	49	47	10	218
Total Conf	65	47	63	73	70	18	336

L'équipe a publié 162 articles de revues et 336 articles de conférences sur la période, ce qui fait une moyenne annuelle d'environ 2 articles en revue et 4 en conférences par permanent. Si le nombre d'articles en revues classées est relativement stable, le nombre d'articles en conférences classées est en augmentation avec une progression de 26% entre 2013 et 2017. La qualité des revues est relativement constante pour les revues de rangs A et A\* cumulées. La proportion d'articles en conférences classées représente environ 65% des articles en conférences et est justifiée par le besoin des chercheurs de participer à des événements scientifiques plus ciblés thématiquement.

### 6.2.1 Bilan scientifique

L'équipe R3S a obtenu de nombreux résultats, en particulier sur les quatre thèmes identifiés comme stratégiques dans le dernier rapport d'évaluation, à savoir la virtualisation des réseaux (SDN) [6.J.rs, 10.J.rs], l'Internet des objets (IoT) [7.J.rs, 4.J.rs, 16.J.rs, 36.C.rs], l'énergie [2.O.rs, 52.C.rs] et les réseaux complexes [21.J.rs, 28.J.rs], les deux premiers thèmes étant au coeur du projet de l'équipe et les deux derniers au coeur du projet du laboratoire.

Ces travaux répondent à un besoin technologique et des enjeux sociétaux actuels. Ils s'appuient sur des outils méthodologiques et de validation scientifiques.

La suite de la section présente un sous-ensemble des travaux les plus significatifs.

#### Réseaux et services

Le groupe RS a obtenu des résultats scientifiques très favorablement accueillis par la communauté. Ces travaux présentés ci-dessous portent sur la virtualisation de services et des réseaux, les architectures de services, l'efficacité des réseaux VANET (Vehicular Ad hoc Network), et les outils et méthodes d'analyse de méta données. D'autres résultats tout aussi intéressants ont été obtenus sur un ordonnanceur pour améliorer le service de voix sur IP dans le réseau LTE (Long-Term Evolution) [16.J.rs], un modèle de données unifié pour une meilleure interopérabilité des services WBAN (Wireless Body Area Network) [23.J.rs], un algorithme de décision pour optimiser l'énergie consommée lors de la collecte de données sur un ensemble de mobiles [22.J.rs], et une solution analytique pour optimiser les ressources et l'efficacité énergétique lors de la diffusion de flux multimédia sur les réseaux cellulaires de type LTE [23.C.rs].

### Une connectivité efficace dans les réseaux VANET

Les applications de sécurité routière sont la principale motivation du développement des réseaux VANETs. Ces applications critiques visent à fournir une assistance dans des situations dangereuses ou difficiles. Elles ont des exigences strictes en termes de délais de bout en bout et de taux de collisions des paquets.

Nous avons proposé trois nouveaux protocoles de contrôle d'accès au support de transmission (protocoles MAC) s'appuyant sur la technique de division temporelle TDMA (Time Division Multiple Access). Le premier protocole totalement distribué s'appuie sur les informations de localisation [36.C.rs], tandis que les deux autres protocoles sont centralisés et utilisent soit l'infrastructure fixe installée aux bords des routes [35.C.rs], soit l'élection dynamique d'un leader par groupe de voitures géographiquement proches qui se charge de la gestion des slots temporels dans son voisinage [50.C.rs]. Par ailleurs, R3S a conçu un protocole de groupage (clustering) qui améliore la stabilité de la topologie du réseau VANET [26.J.rs]. Les résultats de simulations montrent leur efficacité par rapport aux protocoles faisant référence dans la littérature.

[24.J.rs] propose de modéliser le nombre de sauts dans une composante connectée d'un réseau V2V (Vehicle-to-Vehicle). Le nombre de véhicules appartenant à une composante connectée est bien connu, notamment par les résultats d'Eitan Altman et ses collègues. En revanche, ce travail s'est focalisé sur le nombre de sauts lorsqu'on considère le plus court chemin. Cela dépend évidemment du profil de trafic routier. Ce problème a été modélisé par une file d'attente de type G/D/1 dont les arrivées sont dépendantes. Nous avons réussi à reconstruire la loi du processus d'arrivée de la file pour tout profil de trafic de véhicules. Nous avons proposé une résolution de la file explicite dans le cas poissonien et la méthode de résolution dans les autres cas. En lien avec ces travaux, il a aussi été proposé un mécanisme D2D (Device-to-Device) réduisant jusqu'à 50% les délais de transmission dans un réseau VANET [59.C.rs] et une architecture VANET interactive pour des applications à fortes contraintes temporelles (santé, urgences,...) [15.J.rs]. Ces travaux ont été validés par simulation.

Ces travaux sont issus d'une thèse en cotutelle avec l'ENSI de Tunis et de visites de chercheurs égyptiens réalisées sur des bourses d'échange égyptiennes.

### Virtualisation des services et des réseaux

La virtualisation des services et des réseaux est prometteuse pour les entreprises qui veulent dématérialiser leurs serveurs et infrastructures physiques pour diminuer leurs coûts. R3S a apporté des solutions concrètes à plusieurs problèmes encore non résolus liés à l'automatisation, la validation avant déploiement, et le placement de services et fonctions réseaux. [10.J.rs, 30.C.rs] proposent un orchestrateur de service réseaux virtualisés conforme à l'architecture de virtualisation des réseaux préconisée par l'ETSI-NFV. Le framework TOSCA (Topology and Orchestration Specification for Cloud Applications) a été étendu en rajoutant à son modèle de données des notions de réseau (ex : points de connexion et fonctions attenantes) essentielles à l'interconnexion physique des services et ressources virtuelles.

Nous avons aussi élaboré plusieurs algorithmes de placement optimal des services réseaux virtualisés dans des infrastructures d'hébergement. Des algorithmes issus de l'optimisation combinatoire (ex : programmation linéaire en nombres entiers), ont été proposés pour servir de référence. Des algorithmes heuristiques performants (proches de l'optimal en qualité et surtout passant beaucoup mieux à l'échelle) ont permis de traiter, parfois en temps polynomial, des graphes avec des milliers de noeuds et de liens pour placer et chaîner des services réseaux conformément à leurs workflows. Ces algorithmes s'appuient sur une représentation matricielle des topologies réseaux (matrices d'adjacence), sur la programmation dynamique, sur une représentation des graphes par des arbres, sur des treillis (multi-stage graphs), ou traite le problème de chaînage de services réseaux comme des cycles [11.J.rs, 17.J.rs, 9.J.rs]. Nous avons aussi défini dans l'écosystème de la distribution de contenus un nouvel acteur - le fournisseur de cache pour de grands producteurs de contenus comme Google ou Netflix. L'approche consiste à placer dans le réseau de l'opérateur un cache virtuel qui reste la propriété du fournisseur de contenus et donc libère le fournisseur de cache de la problématique des droits d'auteur. Elle permet d'améliorer les performances de 25% sur la réduction des débits, et 30% sur l'amélioration de la qualité d'expérience [12.J.rs, 15.C.rs]. Le test et la validation avant déploiement de ces services et fonctions réseaux ont aussi fait l'objet de travaux de recherche (notamment en collaboration avec l'équipe METHODE) où les membres préconisent des méthodes de test et de validation par les modèles (représentation en EFSM-Extended Finite State Machines), "white" et "black box testing" [20.J.rs, 13.C.rs].

Ces travaux sont le fruit de travaux de recherche hors et sur contrat, avec des soutiens obtenus sur les projets : FUI 18 DVD2C (2015-2018), FUI 23 CARP (2015-2017), CELTIC SENDATE (2016-2019), ITEA3 EASI-CLOUDS (2011-2015) et bilatéraux avec Orange Labs et Alcatel Lucent Bell-Labs (maintenant Nokia). Ces travaux scientifiques ont pu être testés sur la plate-forme Cloud et réseaux financée en partie par l'Equipe FIT (Future Internet Testing) et la région.

## **Architecture de services**

Les travaux menés au sein de R3S portent sur les aspects services de l’IoT, la sofwarisation des services, ainsi que sur l’analyse de données notamment issues de réseaux sociaux. Parmi les nombreux résultats, seuls trois sont donnés en exemple.

Dans le domaine de l’IoT et des réseaux sociaux, les contributions concernent les couches services, la sémantique, l’interfonctionnement, l’utilisation et l’analyse de données IoT et issues de réseaux sociaux utilisant des concepts avancés de Machine Learning et Deep Learning. Par exemple, [13.J.rs] introduit dans ce cadre un nouveau mécanisme pour la découverte des objets les plus appropriés à utiliser dans un contexte donné. Ces travaux se placent dans le paradigme du Social Internet of Things (SIoT), où les liens entre objets sont modélisés à l’instar des liens entre personnes dans un réseau social. Ce graphe social d’objets constitue ainsi un exemple de couche de virtualisation, tout comme pouvait l’être le graphe d’informations de contexte.

Les travaux ont notamment consisté à développer un framework permettant de réaliser cette couche de virtualisation des objets, et des mécanismes d’exposition vers une couche service, sous forme de recommandation des services les plus pertinents à utiliser dans un contexte donné [25.C.rs]. Pour cela, une modélisation de cette couche de virtualisation a été réalisée sous forme d’une ontologie (incluant notamment objets, personnes, contexte, activités, etc.). Des mécanismes d’expression de requêtes ainsi que des règles de correspondances et de filtrage ont été définies en lien avec cette modélisation.

La granularité des composants a fait l’objet de nombreux travaux et reste encore aujourd’hui discutée. Notre contribution réside dans une méthode de définition de cette granularité [6.L.rs, 34.J.rs] avec une méthodologie pour partir de la description textuelle du service global auquel ce composant participe, c’est-à-dire de l’enchaînement des actions à réaliser par les différents acteurs métiers afin de rendre ce service. Le périmètre du composant et ses interfaces sont ensuite déduits de cette activité métier corrélée avec les technologies d’implémentation utilisées. Cette méthode permet également d’évaluer le niveau de couplage dans une architecture, en mesurant la distribution des interfaces issues d’une même activité métier sur des composants divers.

Ces travaux ont été réalisés dans le cadre des projets H2020 reTHINK, CONVINCe (prix Celtic-Plus Excellent award), H2020 Wise-IoT, ITEA CAP, FP7 eCousin (label EU "Excellent"), SEAS (ITEA Award of Excellence for Innovation and Business impact), ITEA FUSE-IT et dans le cadre de collaborations industrielles, notamment avec Orange et Telecom Italia. Ils ont été particulièrement bien disséminés avec [3.O.rs] classé comme l’un des 10 articles les plus populaires d’IEEE Institute en 2016 .

## **Outils et méthodes d’analyse de métadonnées de téléphonie mobile**

Les méta-informations produites par les équipements de télécommunication mobile sont des sources d’information géoréférencée uniques, par leur large couverture du territoire et de la population. Néanmoins, l’exploitation de ces informations qu’elles soient issues de CDR (Call Details Records) ou d’informations de contexte mobile (protocole PDP - Packet Data Protocol) nécessite un traitement complexe afin de pallier à la faible précision de l’information géographique et au bruit spatial. Bien que ce type d’informations soit encore sous exploité par les opérateurs de téléphonie mobile, les opportunités applicatives issues du traitement de ce type de données sont nombreuses, surtout dans le domaine des transports et de planification dynamique des réseaux.

Dans cette thématique de recherche, nous avons conçu des outils d’analyse macroscopique des déplacements urbains dans un contexte de mobilités multimodales, et plus précisément, une méthode de reconnaissance de trajectoires mobiles basée sur des chaînes de Markov cachées [21.J.rs], qui a aussi donné lieu à un dépôt de brevet [4.O.rs]. Cette méthode est la première à notre connaissance à pouvoir faire de l’inférence de trajectoire multimodale (sachant différencier le mode de transport), uniquement basée sur les métadonnées de téléphonie mobile. Dans la même thématique, nous avons développé une méthode dynamique d’estimation de la densité de population en temps réel [31.C.rs] en exploitant la volumétrie du trafic mobile. Cette méthode a été testée lors de la coupe d’Europe de football en France sur la fréquentation des enceintes sportives.

Les méthodes décrites précédemment se concentrent essentiellement sur des problématiques liées au domaine des transports, néanmoins, nous avons aussi développé deux méthodes applicables aux réseaux informatiques. Une première méthode permet de simuler des mécanismes de diffusion de l’information dans les réseaux mobiles à large échelle s’appuyant sur la technologie D2D [28.J.rs]. Cette technique permet de simuler le comportement d’un mécanisme de diffusion à l’échelle d’un pays à partir des informations de déplacement issues des données de téléphonie mobile (données obtenues durant le challenge D4D d’Orange). Enfin, nous avons défini un outil de classification de charge sur les réseaux mobiles qui se base sur une méthode de classification à séparateurs à vaste marge [34.C.rs].

Ces travaux font partie d'un effort de recherche conjoint des équipes R3S, ACMES et ARMEDIA, mais aussi de partenariats directs avec des opérateurs de télécommunication SFR et Bouygues, d'une collaboration avec une équipe CNR de l'Université Polytechnique de Turin, d'un projet de collaboration scientifique du ministère des Affaires étrangères IMHOTEP (2015-2017) et d'un projet FUI 17 FluidTracks (2014-2017).

## Sécurité des systèmes et des réseaux

Le groupe SSR a obtenu des résultats d'intérêt sur la sécurité des réseaux virtualisés (SDN - Software-defined Networking) et la sécurité et la protection des données dans l'IoT, l'IoT étant pris sous l'angle d'un réseau hétérogène fortement contraint ou bien comme un système cyber-physique utile au système SCADA de contrôle-commande dans l'industrie.

### Cybersécurité pour les réseaux SDN

La technologie SDN est fortement appréciée aujourd'hui des fournisseurs d'accès qui profitent de ses capacités d'abstraction afin de réduire leurs coûts d'opération. Cependant, le contrôleur SDN constitue un point individuel de défaillance, vulnérabilité qui a pu freiner l'adoption de cette technologie. Par ailleurs, les promesses d'automatisation requièrent le développement d'applications tirant partie de la programmabilité des équipements réseau offerte par le SDN.

[6.J.rs, 44.C.rs] proposent un framework basé sur la technologie SDN pour lutter contre les attaques par déni de service distribué (DDoS). A notre connaissance, c'est le premier framework générique déployable sur un contrôleur SDN, qui permet d'une part l'intégration de composants de supervision, détection, corrélation et réaction dans la boucle de sécurité, et d'autre part la coopération entre des clients (capables de bien détecter l'attaque mais pas de se défendre) et les fournisseurs d'accès (qui peuvent, en limitant l'attaque, protéger l'ensemble de leurs clients). Le framework a été validé par simulations et sur la plate-forme THD-SEC.

[4.C.rs] propose une approche originale de contrôle d'accès dynamique pour protéger le contrôleur SDN contre les applications malveillantes, et ce pour un coût en latence raisonnable de 0,5%. L'approche repose sur l'application de règles d'accès dynamiques avec mise à jour possible par un système de détection d'intrusions IDS.

Nous avons aussi élaboré une approche formelle nouvelle [24.C.rs] visant à formaliser la migration de topologies virtuelles et à vérifier le maintien des propriétés de sécurité lors d'une migration. Cette contribution s'appuie sur le SDN pour abstraire aux utilisateurs la correspondance entre une topologie virtuelle et les équipements réseaux physiques, et donc leur permettre de reconfigurer leurs topologies.

Ces travaux trouvent place dans les projets collaboratifs FP7 NECOMA (2013-2016) et H2020 SUPERCLOUD (2015-2018). Le projet NECOMA (Nippon-European Cyberdefense-Oriented Multilayer threat Analysis), une collaboration entre l'Europe et le Japon a été coordonné par Hervé Debar, membre de l'équipe R3S.

Ces travaux scientifiques ont pu être testés sur la plate-forme CyberSécurité pour le Très Haut Débit (THD-SEC) gérée par l'équipe R3S et dont le financement provient du projet ASTRE/THD-SEC (2015-2016), financé par une action de soutien à la technologie et à la recherche en Essonne (ASTRE).

### Résilience des systèmes cyber-physiques

L'industrie dite 4.0 vise la numérisation de toute la chaîne de production et en particulier l'ouverture des protocoles industriels dits SCADA (Supervisory Control And Data Acquisition) pour permettre le pilotage à distance du processus physique. Ces protocoles, du fait de l'absence de protection et de leur ouverture, sont rendus très vulnérables aux attaques. C'est tout le système cyber-physique, c'est-à-dire l'ensemble des entités participantes - informatiques et physiques - qui est fragilisé.

Nous nous sommes intéressés à améliorer la résilience des systèmes cyber-physiques en développant des approches permettant d'étendre des détecteurs de failles pour traiter aussi des attaques intentionnelles. L'approche suivie combine des techniques de détection et réaction dans les réseaux informatiques avec des techniques de détection d'erreurs du domaine industriel. Elle enrichit donc le modèle de détection d'anomalies en recouvrant les couches physique et protocolaire des systèmes cyber-physiques [7.J.rs, 8.J.rs, 27.C.rs].

L'efficacité des approches a été validée au travers de maquettes qui ont une dimension originale puisque les systèmes cyber-physiques sont pilotés via des protocoles SCADA depuis un réseau programmable en utilisant la technologie SDN. Ces maquettes ont été réalisées sur la plate-forme THD-SEC de Télécom SudParis. Certaines d'entre elles sont visibles en ligne sous forme de vidéos à l'URL <http://j.mp/TSPScada> et le code Java et Python est disponible sur github à partir des URL publiées sur <http://j.mp/TSPScada>. Notons en particulier l'utilisation des environnements cyber-physiques qui intègrent des stratégies de contrôle pour expérimenter et analyser des

contraintes en temps réel. Les maquettes ont été conçues pour tester aussi des scénarios d'attaque. Le but final étant d'analyser l'efficacité des nouvelles méthodes de sécurité mises en oeuvre sur les environnements émulés.

Ces travaux trouvent place dans le projet collaboratif FP7 PANOPTESSEC (2013-2016) qui a été coordonné par R3S et la chaire Cyber CNI (Cybersécurité des Infrastructures critiques) de l'Institut Mines-Télécom.

### **La sécurité et la protection des données personnelles dans l'IoT**

L'IoT insuffisamment sécurisé, comme l'a prouvé le virus Mirai en octobre 2016, représente un risque majeur pour le fonctionnement d'Internet, la sécurité et la vie privée des utilisateurs. La multiplicité des objets potentiellement en interaction, leur hétérogénéité et leurs capacités très réduites freinent la définition de solutions.

Nous concevons des protocoles ou des fonctions de sécurité, à partir de cryptosystèmes existants ou de problèmes mathématiques réputés difficiles, pour répondre à un besoin pratique, en mettant l'accent sur la légèreté des solutions de sécurité retenues, la robustesse des solutions vis-à-vis de modèles de menaces, les propriétés de sécurité ou de vie privée visées, et la facilité de déploiement. Les propriétés sont validées grâce à des jeux de sécurité sous l'hypothèse de modèles (random oracle, standard), ou bien d'outils de vérification formelles (ex : ProVerif).

[45.C.rs] propose un schéma de signcryption, pour simultanément chiffrer et signer, à partir du schéma de signature standardisée KCDSA (Korean Certificate-based Digital Signature Algorithm). Cette approche basée sur l'identité a l'avantage de ne pas nécessiter de certificats de clés publiques EC (Elliptic Curve) et de remplir plusieurs propriétés de sécurité d'intérêt - comme la confidentialité face à des attaquants internes, la non falsifiabilité, et la vérification publique - et ce à des coûts réduits. Les tests sur une plate-forme Wismote<sup>1</sup> ont démontré qu'un chiffrement prenait 2.5s et consommait 16.25mJ pour un niveau de sécurité 80.

Deux autres approches [4.J.rs, 29.C.rs] ont également été conçues avec l'objectif d'externaliser une partie des calculs de chiffrement sur un serveur considéré de confiance relative (semi-trusted) de sorte que l'objet préserve ses ressources, mais que les données externalisées ne mettent à mal ni la sécurité, ni la vie privée. Pour [4.J.rs] qui externalise une partie des calculs d'un chiffrement ABE (Attribute-Based Encryption), il est possible de vérifier que le calcul opéré par le serveur est correct, aussi de cacher l'arbre d'accès utile au déchiffrement pour une meilleure préservation de la vie privée. Enfin, notons la solution [35.J.rs] qui définit un modèle pour gérer la confiance en prenant en compte le contexte, la capacité des noeuds IoT à remplir certains services, et la réputation des noeuds acquises vis-à-vis de ses pairs.

Ces travaux trouvent place dans la chaire VPIP (Valeurs et politiques des informations personnelles) de l'Institut Mines-Télécom inaugurée en 2013 et deux collaborations avec le CEA sur les périodes 2010-2013 et 2013-2016.

## **6.2.2 Faits marquants de l'équipe R3S**

L'équipe R3S a obtenu plusieurs belles réalisations scientifiques dans la période. Nous avons organisé plusieurs conférences internationales, en particulier la 19<sup>ème</sup> édition du symposium RAID (Research in Attacks, Intrusions and Defenses, rang A dans le référentiel CORE) en 2016 à Evry pour laquelle nous avons tenu le rôle de président en plus de membre du Steering Committee. Nous sommes intervenus comme éditeurs associés de plusieurs journaux de renom comme Computers & Electrical Engineering (rang B ERA) ou Computers & Security (rang B ERA). Nous avons coédité plusieurs numéros spéciaux associés à des journaux de référence comme Security and Computers and Electrical Engineering [4.L.rs], World Wide Web Journal [1.L.rs], et plusieurs ouvrages scientifiques [3.L.rs, 6.L.rs, 2.L.rs]. Nous avons été invités pour plusieurs séjours dans des laboratoires internationaux (NTU, Politecnico di Torino).

Les publications scientifiques de R3S ont été primées dans plusieurs conférences APCC 2017 (23<sup>rd</sup> Asia-Pacific Conference on Communications, Perth, Australia - rang B CORE), et IWSEC 2016 (11<sup>th</sup> International Workshop on Security, Tokyo, Japan - rang B CORE).

Même si les travaux publiés ces 5 dernières années sont relativement jeunes, certains d'entre eux ont eu un impact en termes de nombre de citations avec le survey sur les réseaux de capteurs sans fil [30.J.rs] cité 471 fois depuis 2014 (Google Scholar - août 2018) et l'article [35.J.rs] cité 212 fois depuis 2013 (Google Scholar - août 2018), tous deux en relation avec l'IoT. Notons également les 24.700 téléchargements de chapitres pour l'ouvrage "Evolution of Telecommunication Services - The Convergence of Telecom and Internet : Technologies and Ecosystems" [5.L.rs] depuis 2013, les 8.100 téléchargements pour le livre blanc "Semantic interoperability for the Web of Things" [2.O.rs] depuis 2016<sup>2</sup> et le livre blanc "Towards 5G Software-Defined Ecosystems" [3.O.rs] qui a été classé no 5 des articles

1. La plate-forme Wismote utilisée a les caractéristiques suivantes : un micro-contrôleur 6 MHz MSP430x, 16 kB de RAM, 128 kB de ROM, interface radio IEEE 802.15.4

2. [http://www.researchgate.net/publication/307122744\\_Semantic\\_Interoperability\\_for\\_the\\_Web\\_of\\_Things](http://www.researchgate.net/publication/307122744_Semantic_Interoperability_for_the_Web_of_Things)

les plus populaires d'IEEE Institute en 2016<sup>3</sup> après les articles sur Shannon ou l'implication de Marie Curie dans la première guerre mondiale.

Notre réseau académique et industriel et notre expertise reconnue à l'international nous ont permis de contribuer à des projets collaboratifs de type H2020, FP7, PICS, STIC-AMSUD, FUI, PHC... et même de piloter plusieurs projets H2020 et FP7 (FP7 NECOMA, FP7 PANOPTESSEC, H2020 EUNITY, H2020 CSA dialogue cyber Europe Japon). Nous avons participé à la sélection de projets H2020 et FCT (Portugal). Au niveau national, nos relations privilégiées avec les industriels, nous ont values de contracter 31 bourses CIFRE avec de nombreux partenaires comme Orange, SFR, EDF, Thalès, Nokia et 9 contrats bilatéraux. Notre démarche de valorisation nous a amenés à déposer trois brevets, dont deux à portée européenne.

Nous sommes très actifs dans la diffusion de notre expertise auprès des plus hautes instances de décision internationales. Depuis 2016, nous participons en tant que membre élu au partnership board du PPP ECSO (Public-Private Partnership European Cyber Security Organisation), qui vise à entièrement revoir et développer le marché européen de la cybersécurité pour des garanties de souveraineté. De plus, nous avons participé à des consultations publiques nationales en lien avec le numérique (par exemple sur les implications du règlement e-privacy). Enfin, nous avons été actifs en normalisation en endossant les rôles de vice-président du groupe ETSI Information Security Indicators et de rapporteur et contributeur du comité technique ETSI SmartBAN.

Enfin, un effort tout particulier a porté sur la diffusion de connaissances scientifiques vers le grand public. On retiendra en particulier la participation à plusieurs interviews radio (France Culture, Chérie FM), l'émission télévisée "les soirées docu-débats" sur Public Sénat et deux reportages pour AFP TV et l'émission "Dans les Secret de ..." sur NRJ12. Plusieurs articles vulgarisés ont aussi été diffusés via la presse en ligne "The Conversation" ou l'actualité scientifique de l'INS2I du CNRS.

Nous avons aussi contribué à diffuser les connaissances scientifiques en participant à l'élaboration de MOOCs ("Programmer l'Internet des Objets", "Théorie des files d'attente", "A la découverte des télécommunications", "Supervision de réseaux") et en organisant en 2014 une école thématique "Secure Smart Objects & the Internet of Things" adossée au GDR ASR en partenariat avec le CNAM.

### 6.3 Organisation et vie de l'équipe R3S

L'animation de l'équipe est réalisée au travers de réunions d'équipe et de journées thématiques sur la sécurité et l'énergie, auxquelles les doctorants et post-doctorants ont été invités à présenter leurs travaux. Des séminaires organisés par R3S et ouverts au public ont aussi été proposés, à raison de 6 par an en moyenne. Plusieurs orateurs prestigieux comme Prof. J.J. Quisquater, professeur émérite à l'Université catholique de Louvain, Belgique ou Dr. Richard Waldinger, SRI International, Menlo Park, CA, US sont intervenus. Enfin, une trentaine de chercheurs ont été accueillis par des membres de R3S pour des séjours d'une durée moyenne de 4 mois.

Le responsable d'équipe gère le budget interne (dotation institutionnelle) avec la double exigence d'une répartition équitable entre les membres et du respect des critères d'excellence portés par le laboratoire. Plusieurs actions ont ainsi été soutenues, comme la participation aux conférences, l'organisation de conférences, la participation aux écoles d'été de nos doctorants et les séjours de chercheurs invités.

### 6.4 Analyse SWOT de l'équipe R3S

L'analyse de nos activités, de nos résultats et de notre positionnement conduit à l'auto évaluation suivante :

---

3. <http://theinstitute.ieee.org/ieee-roundup/blogs/blog/the-institutes-10-most-popular-articles-of-2016>



Points forts	<ul style="list-style-type: none"> <li>— Equipe de taille importante avec une expertise large permettant d’aborder de nombreuses problématiques dans les réseaux, expertise qui sera renforcée fin 2018 par l’intégration de trois enseignants-chercheurs de qualité nouvellement recrutés.</li> <li>— Très bon niveau de visibilité nationale et internationale mesuré par le niveau de publications en revues et conférences classées, la prise de responsabilité dans les conférences/workshops et les comités d’expert (H2020, GDR, ANR, IRT Systematic).</li> <li>— Forte croissance des collaborations internationales et nationales mesurées par l’augmentation du nombre et de la qualité des chercheurs invités pour des séjours de plusieurs mois et des productions scientifiques associées.</li> <li>— Engagement de plus en plus important auprès des instances européennes (PPP ECSO) et nationales (Minefi, pôle de compétitivité).</li> <li>— Forte capacité à innover sur des thèmes d’intérêts économiques et sociétaux forts.</li> <li>— Effort de diffusion des recherches très important auprès du grand public.</li> <li>— Essor de collaborations intra et inter équipes et aussi inter laboratoire (LTCl) qui ont abouti à des copublications et des coencadrements de thèses.</li> <li>— Sources de financements très diversifiées - chaires, H2020, ANR, DGE (FUI, PIA), bourses CIFRE et IMT - permettant à l’équipe d’être relativement sereine quant à la capacité de financer ses recherches.</li> <li>— Intégration réussie de l’équipe à l’Université Paris-Saclay des prises de responsabilité au sein de l’école doctorale STIC (adjoint à la direction du pôle 2), dans les Masters informatique et dans le labex Digicosme (activités d’animation).</li> </ul>
Points faibles	<ul style="list-style-type: none"> <li>— Manque d’ingénieur (hors contrat) pour développer des travaux hors contrat. R3S sollicite régulièrement le CNRS pour le financement d’ingénieur plate-forme. Nous avons obtenu le financement d’un ingénieur sur 12 mois par le CNRS pour la plate-forme Cloud, mais cela reste insuffisant pour nos besoins.</li> <li>— Marge de progression possible de collaborations inter et intra équipe, les mécanismes d’incitation devant provenir à la fois de l’équipe et du laboratoire.</li> </ul>
Opportunités	<ul style="list-style-type: none"> <li>— Profiter de l’intégration de l’équipe dans NewUni pour entamer des collaborations de recherche et d’enseignement avec les chercheurs des laboratoires partenaires. Forts de l’expérience avec l’UPSaclay, nous avons l’opportunité de reconstruire nos partenariats de façon plus avisée.</li> <li>— Bénéficier de l’écosystème de NewUni autour des thèmes DATA-IA et sécurité pour renforcer nos compétences et faciliter la réalisation de travaux novateurs.</li> </ul>
Risques	<ul style="list-style-type: none"> <li>— Risque de ralentissement du recrutement d’excellents doctorants et chercheurs, du fait de la complexité administrative introduite par le fonctionnement en parallèle de deux écoles doctorales ED-STIC et ED-NewUni, et des délais nécessaires à la promotion du nouveau label “NewUni”.</li> <li>— Difficulté à intégrer la bilocalisation. Ce risque est faible compte tenu que la bilocalisation est déjà une réalité pour un cinquième des effectifs de R3S.</li> <li>— Risque de tarissement des financements dans un contexte économique tendu où les flux financiers sont peu nombreux et très compétitifs, ce qui pourrait avoir pour conséquences une dépendance trop forte vis-à-vis des industriels et des travaux de recherche beaucoup moins théoriques. Le risque est relativement faible à moyen terme du fait de la diversification des financements amorcée par l’équipe.</li> </ul>

## 6.5 Projet scientifique de l’équipe R3S

Les projets portés par R3S sur les cinq prochaines années s’inscrivent dans la démarche de numérisation et de virtualisation de tous les services. Cette démarche a démarré depuis quelques années et va en s’amplifiant. Elle touche différents secteurs des services publics et de l’industrie tels que la santé, l’énergie, mais aussi toute la chaîne de production industrielle. Elle implique de s’appuyer sur des systèmes communicants ramifiés et efficaces dans la

gestion des interactions et sur des architectures distribuées pour assurer un service de qualité et une administration performante du service.

Forts de notre expertise dans les systèmes et objets communicants et la virtualisation, nous voulons nous impliquer sur les deux projets décrits ci-dessous, à savoir l'optimisation de ressources pour des services sous contraintes, et la sécurité et protection des données personnelles dans le cas d'applications critiques touchant à la santé et à la production industrielle. L'équipe veut de la sorte participer au développement économique du numérique, et satisfaire le délicat équilibre entre une bonne qualité d'expérience pour l'utilisateur, une consommation de ressources réduite pour limiter l'empreinte carbone du service résultant, le respect des règles législatives en matière de traitement des données personnelles et la protection des personnes et des intérêts économiques des Etats et des industriels.

### 6.5.1 Optimisation de ressources pour des services sous contraintes

Ce projet répond au problème de déploiement de services sous contraintes et vise à ajuster de façon continue, dynamique et rétroactive le déploiement de services pour satisfaire au mieux un ensemble de contraintes. Les contraintes peuvent porter sur des ressources consommables, auquel cas elles évoluent dynamiquement en fonction des ressources disponibles restantes, mais elles peuvent aussi porter sur un ensemble d'exigences en termes de sécurité et du niveau de confiance entre entités. La notion de service, quant à elle, est très large puisqu'elle couvre les services virtualisés ou non, et/ou les services directement consommés par des utilisateurs ou répondant à des exigences de fonctionnement.

Le projet s'intéresse à ce problème, sous les angles algorithmique et d'analyse statistique pour exploiter les gros volumes de données collectées et adapter le déploiement et la gestion du cycle de vie des services. Plusieurs défis se posent tels que le passage à l'échelle des techniques usuelles de traitement de données, la prise en compte de l'hétérogénéité des données et des technologies supports (Ethernet, optique, sans fil), et la présence d'un environnement multicouches (notions d'"overlays" multiples).

Trois cas d'usage sont envisagés :

- L'optimisation des services virtualisés, sous contraintes de qualité d'expérience (QoE) et de ressources : Pour les services et fonctions réseaux virtualisés, la production et l'adaptation automatique de services dans un contexte multi-locataires (tenants) doit s'appuyer sur l'analyse des données de monitoring des ressources et le contrôle et la programmabilité des infrastructures d'hébergement. L'objectif est de prédire la demande, d'anticiper les congestions, de détecter les anomalies et dégradations, d'identifier les causes racines pour maintenir la qualité de service et d'expérience. Les systèmes actuels, confrontés à des environnements de plus en plus stochastiques, doivent combiner science des données, science des réseaux et optimisation de systèmes dynamiques pour répondre au besoin.

Pour le slicing d'infrastructures partagées pour les réseaux du futur (5G, IoT et cloud), faisant suite à nos travaux [17.J.rs, 11.J.rs], nous comptons adresser deux verrous. Le premier verrou est la prise en compte de l'accès radio, et des contraintes qu'il impose à l'orchestration et au déploiement de service de bout en bout. Une piste consiste à modéliser et caractériser les aspects trafic, spatio-temporels et localisation dans les réseaux 5G grâce à la géométrie stochastique, de considérer les passerelles en bordure des réseaux radio comme des racines d'arbres d'où proviennent les flux et demandes vers le réseau cœur (ou xHaul), et de résoudre le problème d'optimisation des ressources de bout en bout, sur la partie cœur, comme un problème d'arbre recouvrant ("spanning tree"). Le deuxième verrou concerne la prédiction fiable de la demande, des changements dynamiques, des dégradations et des intrusions dans le système. Il s'agit de comprendre les corrélations et les dépendances entre les changements dynamiques et les ajustements de ressources, de sorte à éviter de fréquentes adaptations des systèmes, pour répondre efficacement aux dégradations et maintenir la qualité de service. Il s'agit aussi de corréler la qualité mesurée dans les réseaux (QoS) et la qualité d'expérience (QoE) pour prédire les dégradations de QoE à partir des mesures de QoS sur le réseau. Les méthodes envisagées pour faire ce lien comprennent les réseaux de neurones (notamment récurrents) et l'apprentissage par renforcement.

- L'optimisation du déploiement et du maintien de la sécurité : Dans le cadre de la fourniture de services dans les réseaux 5G, qui nécessite la mise en œuvre de réseaux virtuels sur des infrastructures multi-fournisseurs et multi-locataires (tenants), l'opérateur de réseaux virtuels doit pouvoir déployer des fonctions de sécurité au plus près de ces services. Ce réseau virtuel doit être résilient, et s'adapter au moindre incident ou changement survenu sur l'infrastructure, ce qui peut aboutir à la décision de migrer le réseau virtuel et les fonctions de

sécurité. Nous envisageons de nous appuyer sur nos travaux antérieurs [24.C.rs] et d'utiliser des fonctions de supervision de l'état du réseau et du niveau de sécurité de celui-ci pour réagir à ces changements. Il est prévu que ces fonctions servent notamment d'entrée à (i) des mécanismes de déploiement dynamique de fonctions de sécurité, afin d'assurer une couverture optimale du réseau, un langage de haut niveau d'abstraction pouvant alors servir à décliner la politique de sécurité en fonction du type d'infrastructure et du cas d'usage ; (ii) des algorithmes de calcul de chemins sécurisés, afin de satisfaire les besoins de sécurité des locataires tout en respectant les contraintes des fournisseurs d'infrastructure, notamment la confidentialité de leur topologie ; (iii) des mécanismes de migration du réseau virtuel permettant de conserver les propriétés de sécurité et de réagir en cas de violation. Ces trois directions de recherche concourent à fournir une couverture de sécurité optimale et en profondeur.

- L'optimisation des réseaux électriques : Dans les réseaux électriques pour lesquels les surplus d'énergie produite peuvent difficilement être stockés, il est nécessaire de gérer finement la consommation et la production d'énergie. Les travaux porteront sur l'optimisation de la gestion de l'énergie, grâce à des données de consommation et de production obtenues d'environnements réels. Nous envisageons de proposer des algorithmes s'appuyant à la fois sur des modèles et des données. Les modèles seront établis en collaboration avec des chercheurs de Samovar et des spécialistes des laboratoires CNRS du GeePS et du LMD. Les outils pressentis viennent de la théorie des files d'attente ou des systèmes complexes (modèles de grands graphes) . Les algorithmes envisagés s'appuieront par exemple sur des approches de réseaux de neurones et sur du renforcement pour permettre aux algorithmes de s'adapter à l'état du système. Une des difficultés du sous-projet réside dans la nécessité d'avoir une connaissance précise du comportement des consommateurs et de la production. Pour cela, il est prévu l'accès à des données de consommation résidentielles et tertiaires ainsi qu'à des données de production photovoltaïque et éoliennes extrêmement précises. Un des enjeux de ces recherches sera de collecter, filtrer, traiter ces données pour obtenir des indicateurs facilement exploitables et surtout de les recouper pour en inférer un maximum d'informations. Par ailleurs, les recherches menées dans le laboratoire sur la mobilité permettront aussi d'investiguer l'impact de la mobilité sur les comportements de consommation et du stockage dynamique d'énergie, en vue d'améliorer la prédictibilité des indicateurs.

## 6.5.2 Sécurité et protection des données personnelles dans le cas d'applications critiques

Les applications critiques considérées ici sont des applications dont le dysfonctionnement intentionnel ou non peut avoir de graves conséquences sur la sécurité des citoyens, l'environnement ou certains services étatiques ou économiques essentiels à la nation<sup>4</sup>. Ces applications ont un besoin crucial d'être protégées contre les cyberattaques, mais elles ont aussi besoin de se conformer au règlement général sur la protection des données personnelles (RGPD)<sup>5</sup> mis en application le 25 mai 2018, et à la directive NIS<sup>6</sup> transplantée dans le droit national par les lois de programmation militaire 2014-2019 et 2019-2025, en attendant une future réglementation européenne actuellement en cours de définition. Dans ce cadre, ce projet s'intéresse à deux types d'applications critiques faisant appel à des systèmes communicants et à des environnements plus ou moins ouverts et dynamiques :

- E-santé, les applications de santé reposant sur des dispositifs médicaux autonomes<sup>7</sup>. Certaines applications, comme les défibrillateurs cardiaques télésurveillés ou les pacemakers autonomes, sont d'importance vitale pour les patients et nécessitent des mesures de sécurité fortes. Elles traitent des données de santé qualifiées de "sensibles" par le RGPD, la divulgation de ces données, y compris la simple information d'être porteur d'un dispositif, pouvant nuire à la qualité de vie du patient par le refus d'un prêt bancaire, d'une assurance ou d'une embauche.
- L'industrie 4.0 qui s'appuie sur les technologies de l'Internet des objets et les systèmes cyber-physiques

4. Code de la défense, Article R.1332-2, Légifrance. <https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006071307>

5. Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (Règlement général sur la protection des données), JOUE L 119/1 du 4 mai 2013, <http://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32016R0679&from=FR>

6. <https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive>

7. Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE

pour numériser toute la chaîne de production industrielle et permettre ainsi une communication continue et instantanée entre différents dispositifs, ceci pour augmenter l'efficacité des processus rendus plus autonomes et améliorer la productivité par l'adaptation des processus et la réduction des coûts humains et énergétiques. Certaines applications comme la distribution d'énergie ou d'eau, la production d'équipements militaires ou d'uniformes personnalisés sont critiques et ne souffrent d'aucune coupure de service. Les données afférant à la production peuvent être sensibles car porteuses par exemple d'informations stratégiques sur le système de défense d'un Etat ou bien de secrets de production industriels.

Ces deux cas d'usage font appel à des systèmes communicants tels que des actionneurs, des capteurs et des contrôleurs, souvent en association avec des services cloud. Certains besoins leur sont communs comme la bonne résistance aux attaques, la résilience du système, la sûreté de fonctionnement, la conformité au RGPD, l'interopérabilité des dispositifs, et la personnalisation du service. Pour l'e-santé, du fait de l'usage à la fois grand public et par le personnel de santé, il est important de veiller à l'ergonomie, l'automatisation et la traçabilité des opérations lors de la mise en service des dispositifs, de l'accès aux données et de l'accès à distance aux dispositifs médicaux, et ce pour répondre aux enjeux actuels et favoriser l'essor de ce secteur économique. Pour l'industrie 4.0, un autre aspect important est à prendre en compte, le besoin pour l'industriel de protéger sa propriété intellectuelle car elle constitue une grande partie de sa richesse économique.

Les techniques et mécanismes suivants sont envisagés pour répondre à ces besoins : 1) l'apprentissage profond pour détecter des anomalies, et pouvoir distinguer une faute d'une attaque afin d'ajuster la politique en réaction, 2) le déploiement de fonctions de surveillance et de fonctions de réactions aux attaques compatibles avec l'hétérogénéité des objets connectés et la rareté des ressources, 3) le couplage fort entre les moyens de sécurisation des protocoles classiques et les mécanismes de gestion de la confiance pour réduire l'impact d'une attaque voire isoler un attaquant, 4) la programmabilité des réseaux SDN pour définir une partition logique du réseau en vue d'isoler un attaquant, 5) la définition d'un mode dégradé permettant à un service de continuer à fonctionner en cas de cyber-attaques en conciliant au mieux les exigences de sécurité et de sûreté de fonctionnement, 6) l'alignement d'ontologies de référence pour gérer l'hétérogénéité des systèmes communicants et l'hétérogénéité inter domaines d'application dans la mise en oeuvre de politiques de préservation de la vie privée, 7) les mécanismes embarqués de type "semantic analytics" à base de règles sémantiques et de raisonnement pour gérer l'hétérogénéité localement au dispositif, 8) des fonctions cryptographiques légères permettant de personnaliser un service tout en limitant la fuite de données vers le service, et 9) la gestion dynamique du consentement en conformité avec le RGPD sous forme de preuves cryptographiques. Ces travaux s'appuieront sur le savoir-faire de l'équipe acquis au travers des résultats de recherches [6.J.rs, 8.J.rs, 27.C.rs, 35.J.rs, 23.J.rs, 2.O.rs, 4.J.rs, 29.C.rs] et sur des collaborations avec les laboratoires académiques (LTCI, LIX) et industriels (EDF, Renault).

# 7 Équipe TIPIC

## 7.1 Présentation de l'équipe TIPIC

### 7.1.1 Introduction

Le périmètre de l'équipe TIPIC (Traitement de l'Information Pour l'Image et les Communications) a changé au cours du quinquennat.

Au 1<sup>er</sup> janvier 2013, les domaines de compétences de l'équipe se situaient à l'interface de la "couche physique" (transmission et traitement de l'information) des STIC et des mathématiques appliquées. L'équipe était structurée en quatre axes principaux : Inférence et restauration bayésienne dans des modèles markoviens ; Traitements statistiques de signaux multi-sources et multi-capteurs ; Algorithmes rapides pour l'électromagnétisme, et Apprentissage statistique et reconnaissance des formes.

En 2015, l'intégration d'enseignants-chercheurs (EC) de l'ENSIIEE, d'une part, et d'EC de TSP faisant jusque là partie de l'UMR CNRS MAP5, d'autre part, a entraîné une restructuration des équipes existantes de SAMOVAR ; en ce qui concerne TIPIC :

- les membres de TIPIC qui faisaient partie de l'axe "Apprentissage statistique et reconnaissance des formes" (B. Dorizzi, S. Garcia et D. Petrovska-Delacrétaz) ont rejoint la nouvelle équipe ARMEDIA ;
- les EC du groupe "technologies avancées pour les communications" (B.-E. Benkelfat, Y. Frignac, Y. Gottesman, C. Lepers, N. Samama, A. Vervisch-Picois, Q. Zou) de l'équipe R3S ont rejoint TIPIC.

Notons enfin que Y. Petetin a rejoint TIPIC en 2015, et D. Clark et T. Taillandier-Loize en 2017.

### 7.1.2 Effectifs et moyens de l'équipe

L'équipe est constituée de 18 membres permanents, dont 10 professeurs, et 8 Maîtres de conférences (dont deux HDR). 2 EC ont soutenu leur HDR dans la période : Y. Gottesman en 2014 et Y. Frignac en 2016.

**Professeurs :** Badr-Eddine Benkelfat, Jean-Pierre Delmas (émérite), François Desbouvries, Randal Douc, Yann Frignac, Frédéric Lehmann, Catherine Lepers, Christine Letrou (à 50%), Wojciech Pieczynski, Nel Samama.

**Maîtres de conférences** Marc Castella, Daniel Clark (depuis 2017), Yaneck Gottesman (HDR), Emmanuel Monfrini, Yohan Petetin (depuis 2015), Thierry Taillandier-Loize (depuis 2017), Alexandre Vervisch-Picois, Qin Zou (HDR).

Par ailleurs, sur la période l'équipe TIPIC a accueilli 44 doctorants (20 par an en moyenne), 5 postdocs et 6 ingénieurs de recherche (certains d'entre eux pouvant d'ailleurs être considérés comme des postdocs). La durée moyenne des thèses est de 39,7 mois, celle des postdocs de 18 mois, celle des ingénieurs de recherche de 13 mois. Le tableau ci-dessous présente l'évolution des personnels permanent et non permanent de TIPIC sur la période 2013-2018.

	1/1/2013	1/1/2014	1/1/2015	1/1/2016	1/1/2017	1/1/2018
permanents	15	15	16	16	18	18
doctorants	17	22	23	24	18	16
post-doctorants	1	1	1	0	3	1
ingénieurs de recherche	1	1	1	3	1	0

Le financement des 44 doctorants se répartit de la façon suivante : 13 bourses CIFRE, 9 financements institutionnels (IMT-TSP, Carnot...), 6 financements sur contrats, 9 externes (salariés en France hors IMT, ou encadrés partiellement par un EC de TIPIC et inscrit ailleurs en France), 3 cotutelles, 4 étrangers. Le tableau ci-dessous précise la répartition des thésards au 1<sup>er</sup> janvier de chaque année, par type de financement :

	1/1/2013	1/1/2014	1/1/2015	1/1/2016	1/1/2017	1/1/2018
CIFRE	3	6	8	9	6	6
IMT-TSP	4	5	5	5	3	2
Contrat	3	3	3	3	2	2
Externe	4	4	3	4	4	4
Cotutelle	1	1	2	2	2	2
Étranger	2	3	2	1	1	0

L'équipe a bénéficié sur la période 2013-2017 des financements suivants (en kE) :

	2013	2014	2015	2016	2017
Programmes européens hors ERC et hors fonds structurels	185	96	356	77	0
Appels à projet ANR (hors PIA)	73	107	42	112	97
Programme Investissement d'Avenir	7	0	16	0	0
Autres financements publics sur appels à projets	28	117	61	53	71
Fondations, associations, mécénats	0	20	106	36	26
Contrats de recherche industriels	94	80	100	171	169
Prestations (expertise, service, conseil, recette de colloque)	0	0	0	20	20
Institut Carnot	7	35	34	0	11
Valorisation des résultats de la recherche	0	0	0	30	0
<b>TOTAL</b>	<b>395</b>	<b>455</b>	<b>715</b>	<b>498</b>	<b>394</b>

Ce tableau inclut les financements complémentaires aux bourses de thèse (et en particulier aux bourses CIFRE), mais non les bourses elles-mêmes (rappelons que sur la période nous avons eu 13 thésards CIFRE et 9 bourses institutionnelles IMT).

### 7.1.3 Profils d'activité

En regroupant les EC en traitement statistique du signal et en technologies avancées pour les communications, la nouvelle équipe TIPIC fédère les EC du laboratoire qui travaillent sur la couche physique (transmission et traitement de l'information) des systèmes de communication. Ce chapeau commun constitue une identité partagée qui positionne les membres de l'équipe par rapport à un laboratoire dont le centre de gravité est plus orienté vers l'informatique et les réseaux.

Les domaines de compétences de l'équipe se situent désormais à l'interface de la physique (technologies optiques pour les réseaux de communication), du traitement statistique du signal (transmission et traitement de l'information, localisation indoor) et des mathématiques appliquées (séries temporelles, inférence bayésienne). Dans ce contexte, les techniques de traitement du signal constituent une passerelle naturelle entre les différents EC du groupe, et sont potentiellement vecteurs de synergies, en particulier dans le domaine porteur des communications sur réseaux optiques : même s'il existe une forte expertise "signal" dans les systèmes de communication sans fil (techniques multi-porteuses T/F/CDMA, OFDM, OFDMA, traitements MIMO...), leur adaptation au domaine des communications optiques n'est pas immédiate car elle nécessite la prise en compte des propriétés spécifiques du support, et bénéficie de l'exploitation de caractéristiques propres telles que la polarisation ou la propagation multi-modale. La nouvelle équipe TIPIC a permis de jouer un rôle dans ce domaine puisqu'elle fait se rencontrer des compétences complémentaires.

Globalement notre activité de recherche se répartit en :

- activité scientifique (production de la connaissance) : 45% ;
- activité de valorisation et de transfert : 25% ;
- appui à la communauté (animation scientifique, GDR, organisation de conférences, pilotage stratégique, participation à des comités, commissions...) : 15% ;
- formation par la recherche : 15%.

## 7.2 Produits de la recherche et activités de la recherche de l'équipe TIPIC

La production scientifique de l'équipe sur la période 2013-2018 est de 88 articles de revues, 121 articles de conférences, une monographie de recherche (publiée chez Chapman et Hall), deux chapitres de livres (publiés chez Springer et Elsevier) et 6 brevets. Sur le plan qualitatif, 19 des 88 articles de revues sont parus dans des revues classées A\* dans la base de données ERA / CORE (qui n'est plus maintenue à jour depuis 2010), 32 dans des revues classées A, et 11 dans des revues classées B.

### 7.2.1 Bilan scientifique

Les activités de recherche de l'équipe TIPIC sont structurées en quatre axes : Estimation dans des modèles markoviens ; Méthodes d'estimation et de détection pour les traitements multicateurs ; Optique et photonique ; et Simulations, propagation et navigation.

#### Estimation dans des modèles markoviens

(D. Clark (depuis 2017), F. Desbouvries, R. Douc, E. Monfrini, Y. Petetin (depuis 2015), W. Pieczynski).

Une des activités de l'équipe TIPIC consiste à développer et/ou analyser des outils généraux de statistique computationnelle (essentiellement basés sur des tirages Monte Carlo) dans le cadre de modèles Markoviens, que ce soient des chaînes de Markov cachés (HMC), qui ont de nombreuses applications en biostatistique, traitement du signal, reconnaissance de la parole, poursuite de cibles, séries financières..., ou des extensions de ces modèles. Les recherches menées dans ce cadre visent à améliorer les algorithmes existants d'estimation bayésienne ; à en analyser les propriétés statistiques ; ou encore à proposer des modèles probabilistes Markoviens qui modélisent mieux les interactions entre les différentes variables du problème, ou dans lesquels les algorithmes d'estimation bayésienne sont plus aisés à mettre en oeuvre.

- **Les algorithmes de Monte Carlo par Chaîne de Markov (MCMC)** estiment un moment d'une distribution cible à partir d'une trajectoire d'une chaîne de Markov approchant cette distribution. Pour pouvoir comparer différents algorithmes MCMC, il est crucial de disposer d'outils d'analyse basés sur la variance asymptotique de l'estimateur. Dans ce cadre, nous avons proposé des outils généraux [15.J.tp] qui permettent d'analyser des algorithmes de type Metropolis-Hastings, mais s'appliquent également à des chaînes possiblement non-réversibles, à des algorithmes à augmentation de données et ont aussi permis de proposer des variantes efficaces exactes sur des algorithmes de type pseudo-marginaux.

- **Les algorithmes de Monte Carlo séquentiels (SMC)** sont une autre classe importante d'algorithmes de type Monte Carlo ; nos apports dans ce domaine sont méthodologiques et statistiques.

Les algorithmes SMC propagent dans le temps une approximation discrète de la loi de probabilité a posteriori de l'état caché sachant l'historique des observations. Ces algorithmes sont basés sur l'utilisation récursive d'une technique d'échantillonnage d'importance normalisé, assorti d'un mécanisme de rééchantillonnage conçu pour limiter la dégénérescence des poids d'importance. Ce mécanisme crée cependant des échantillons dépendants et rétrécit le support de la loi empirique, aussi s'avère-t-il inefficace dans des situations sévères. Nous avons proposé dans ce contexte un mécanisme de régénération basé sur du rééchantillonnage indépendant [5.J.tp]. Par ailleurs nous avons proposé de modifier le mécanisme d'échantillonnage d'importance normalisé de façon à mieux prendre en compte la région cible d'intérêt [1.C.tp].

En termes d'analyse, nous avons essentiellement deux résultats. D'une part, nous avons proposé une analyse comparative locale (en temps) mais non asymptotique (en nombre d'échantillons) de deux algorithmes proches, l'algorithme de filtrage particulaire avec loi d'importance optimale, et l'algorithme de filtrage particulaire auxiliaire totalement adapté [19.J.tp] ; d'autre part, il est important de vérifier que l'erreur particulaire d'approximation ne dégénère pas au fur et à mesure des itérations successives de l'algorithme, ce qui permet de maintenir une taille de population contrôlée tout en avançant dans les itérations temporelles. Cette non-dégénérescence était déjà constatée pour des espaces d'états finis ou compacts, mais l'étendre à des espaces possiblement non compacts (ce qui est souvent le cas en pratique) représentait un défi théorique. Nous avons montré [16.J.tp] que sous des conditions très générales, la variance et plus généralement

les normes  $L_p$  sont bornées en probabilité indépendamment des itérations de l'algorithme, ce qui justifie a posteriori l'utilisation de ces algorithmes particuliers à nombre fixe de particules.

- Une autre classe d'algorithmes d'estimation récursive est constituée des algorithmes de **filtrage multi-cibles**, qui consistent à estimer les paramètres cinématiques (position, vitesse...) d'un ensemble de cibles dont le nombre est aléatoire et évolue dans le temps. Notre collègue Daniel Clark, embauché en 2017 (un des leaders mondiaux dans ce domaine) a développé le premier algorithme multi-cibles [1.J.tp] qui propage les moments des deux premiers ordres, et qui est donc l'analogie multi-cibles du filtre de Kalman. Par ailleurs, l'utilisation pratique de ces algorithmes sur données réelles nécessite l'estimation des paramètres inconnus des modèles sous-jacents. L'article [3.C.tp], par exemple, estime conjointement, dans le cadre d'une application biomédicale, le nombre et la position d'un ensemble d'objets ainsi que le taux de fouillis.
- Un **modèle de Markov partiellement observé (POMM)** est une chaîne de Markov bivariée  $(X_k, Y_k)_{k \geq 0}$  où seul le processus  $(Y_k)_{k \geq 0}$  est observé. Ces modèles englobent entre autres i) les chaînes de Markov cachées (HMC), et ii) les modèles dirigés par les observations, particulièrement utilisés en économétrie (les modèles GARCH en constituent un modèle phare). Bien que ces modèles s'inscrivent tous deux dans le cadre général des POMM, ils ont jusqu'à présent été étudiés séparément, sans doute parce que dans les modèles dirigés par les observations, le nouvel état caché est une fonction déterministe de l'état et de l'observation précédents, ce qui rend impropre la notion d'estimation de l'état caché qui est un problème d'intérêt majeur dans les modèles HMC (voir ci-dessus). Néanmoins des études récentes montrent que pour ces deux sous-classes de modèles, certaines propriétés essentielles fréquentistes (comme le Maximum de Vraisemblance [12.J.tp]) ou bayésiennes (comme l'estimateur a posteriori) peuvent s'analyser dans le cadre commun des POMM, ce qui a nous permis de proposer une théorie de l'estimation unifiée pour ces modèles plus généraux.
- Les POMM (ou modèles couple) ci-dessus se généralisent en des modèles de Markov triplet, c'est-à-dire des séries temporelles dans lesquelles trois processus (l'état caché  $X_k$ , l'observation  $Y_k$  et un processus auxiliaire  $U_k$ ) forment conjointement une chaîne de Markov. L'introduction de modèles triplet dans lesquels le processus caché est continu a ouvert la voie à la conception de nombreux modèles originaux permettant des traitements au premier ordre (que ce soit pour le filtrage [6.J.tp] [4.C.tp], le lissage [13.C.tp] ou la prédiction([5.C.tp]) exacts et rapides. Un des intérêts de tels modèles est de pouvoir approcher tout système  $(X_k, Y_k)_{k \geq 0}$  Markovien possiblement non linéaire et/ou non Gaussien ; cette **nouvelle classe de méthodes bayésiennes variationnelles** constitue donc une alternative aux méthodes SMC.
- Des recherches ont également été menées autour du **pouvoir modélisant des modèles couple et triplet** [2.J.tp] et de leur apport en termes d'applications [7.C.tp]. Ces études, en particulier, ont mis en évidence la robustesse des modèles triplet pour la segmentation non supervisée d'images texturées de l'oeil partiellement altérées, ainsi qu'en imagerie astronomique dans le cadre de la segmentation d'images hyperspectrales de champs profonds obtenues grâce au nouvel outil MUSE du Très Grand Telescope (VLT) européen [7.C.tp]. Nous avons également proposé un ensemble de modélisations originales répondant à des besoins applicatifs en génétique [18.J.tp] et en astronomie [3.J.tp].
- Les études des modèles de Markov triplet permettant l'**utilisation de la théorie de l'évidence dans un contexte Markovien** ont permis de proposer une modélisation générale pour la fusion des données, utilisable aussi bien dans le cas des chaînes [13.J.tp] que des champs. En particulier, l'utilisation de champs de Markov triplet pour modéliser la non stationnarité au niveau du bruit plutôt qu'au niveau de l'image des classes a donné des résultats encourageants [11.C.tp].

## Méthodes d'estimation et de détection pour les traitements multicapteurs

(M. Castella, J.-P. Delmas, F. Lehmann).

L'apparition de nouvelles applications comme les réseaux de capteurs (ex : l'internet des objets), les communications dans des bandes de fréquence exotiques (ex : 5G) ou le traitement de données en grande dimension (ex : Big data) motivent la conception et l'analyse de nouveaux algorithmes permettant d'exploiter efficacement le degré de liberté supplémentaire qu'est la dimension spatiale de ces signaux. Afin de répondre à ces nouveaux défis, nous avons proposé de nouveaux modèles, méthodes statistiques et techniques d'optimisation.

- Le développement du concept d'internet des objets prévoit qu'un grand nombre d'objets de la vie courante (appareils ménagers, automobiles, vêtements, etc) seront dans un futur proche connectés sans fil. Ce nouveau paradigme nécessite le développement de techniques d'accès innovantes permettant de limiter considérablement le nombre de données d'apprentissage et de transmettre simultanément les données en provenance



de plusieurs objets. Dans ce contexte difficile de mélange de signaux superposés, nous avons développé des **modèles probabilistes mixtes discrets/continus** permettant le traitement statistique multicapteur pour l'estimation conjointe des canaux de transmission et des données transmises [20.J.tp]. D'autre part, la simplification à des systèmes mono-antenne pose un défi en termes d'annulation d'interférence que la modélisation explicite des dépendances temporelles permet de résoudre de façon quasi-optimale [11.J.tp].

- La densification des réseaux de communication mobile et les besoins de débits croissants nécessitent une utilisation toujours plus efficace d'une ressource spectrale très limitée. Nous avons exploré la solution consistant à prendre en compte **la noncircularité et la cyclostationnarité** des signaux de communication. En particulier, nous avons quantifié les performances en réjection d'interférences de divers beamformers spatiaux en présence d'interférences et avons démontré la rupture existant entre interférences rectilignes et interférences noncirculaires nonrectilignes [4.J.tp]. De même, nous avons démontré une perte de performance en présence d'interférences quasirectilignes par rapport à des interférences rectilignes, contrairement à ce qui était considéré jusqu'alors dans la littérature. En exploitant les différences de spectres cycliques de ces deux types d'interférences, nous avons proposé des traitements de type maximum de vraisemblance d'estimation de séquence permettant d'atténuer cette perte de performances [17.J.tp].
- Les problèmes de reconstruction de signaux se ramènent souvent en un problème d'**optimisation d'un critère**, dont la valeur est minimale pour les vraies données. D'une part, dans un contexte de séparation de sources mélangées sur plusieurs capteurs, nous avons étudié les problèmes de maximisation de fonctions de contraste et justifié certains schémas itératifs d'optimisation [2.L.tp]. D'autre part, nous avons considéré des critères dans la classe des fonctions rationnelles. Pour la première fois dans la communauté, nous avons fait appel à des méthodes d'optimisation récentes, qui consistent en une hiérarchie de relaxations convexes du problème rationnel de départ [15.C.tp, 8.C.tp]. La richesse de la modélisation rationnelle a permis de dépasser des hypothèses souvent limitantes (modèle linéaire, a priori de parcimonie convexe), tout en offrant des garanties théoriques d'optimalité globale et non plus uniquement locale.

## Optique et photonique

(B.-E. Benkelfat, Y. Frignac, Y. Gottesman, C. Lepers, Q. Zou).

Les activités en optique et photonique trouvent leurs applications dans le domaine du numérique selon deux axes : le transport de l'information et le traitement des données associé ; les capteurs et l'imagerie non-conventionnelle. L'objectif des travaux qui sont menés dans le domaine du transport de l'information est d'une part de répondre à la demande croissante en télécommunication mondiale, induite par les nouveaux modèles numériques appliqués à tous les secteurs industriels (smart-city, smart-grid ou smart-building, etc.) et les grandes quantités de données générées par l'internet des objets et la 5G, et de développer les technologies émergentes des data-centers, de l'accès optique, filaire ou en espace libre, tout en limitant leur coût et leur consommation énergétique d'autre part. Ceux qui sont menés dans le domaine des capteurs photoniques portent aussi bien sur la conception et la réalisation de nouveaux capteurs et imageurs que sur le développement des méthodes d'exploitation des données délivrés par ceux-ci ; les domaines applicatifs visés sont ceux des capteurs biométriques hautement sécurisés, et des outils au service de la santé et des sciences du vivant. Nous avons plus précisément étudié les points suivants :

- **Transport de l'information : fonctionnalités pour les architectures optiques flexibles (flex-rate, flex-grid,...) et efficaces en énergie.** Les communications optiques ont largement évolué depuis leur apparition au milieu des années 90. Parallèlement, les réseaux de transport optiques longue distance sont devenus des réseaux maillés, permettant une reconfiguration des liens entre les noeuds de routage grâce à des multiplexeurs d'insertion/extraction reconfigurables (ROADM). La couche optique est devenue une couche intelligente, capable d'établir des canaux WDM à la demande, de maintenir l'intégrité des connexions en place, ou d'opérer une restauration rapide des canaux WDM en cas de défaillance d'un équipement ou d'un lien. L'augmentation du débit dans les réseaux de transport optiques conduit les opérateurs à déployer des réseaux à 100 Gbps par canal voire 400 Gbps, basés sur le concept d'"élasticité". Dans le contexte de ces réseaux optiques flexibles, nous avons étudié de nouveaux outils de dimensionnement [16.C.tp], considéré le problème de la fragmentation du spectre WDM flexible [19.C.tp], évalué un cas d'usage alliant élasticité et restauration optique [9.J.tp, 12.C.tp] et proposé une solution au problème lié à l'utilisation des amplificateurs optiques conventionnels dans les réseaux flexibles [10.J.tp].
- **Transport de l'information : accroissement du débit des transmissions sur fibre optique longue distance.** La recherche de l'accroissement du débit d'information transmis par les systèmes sur fibre optique

viser l'augmentation de la densité spectrale d'information et de la bande passante optique.

Pour le premier point, des études ont été menées pour les transmissions cohérentes multiplexées en polarisation (gain d'un facteur 2) notamment pour la compréhension des effets non-linéaires de type Kerr entre polarisations, la recherche de formats de modulation résistants à ces effets et le traitement du signal adapté après réception cohérente [22.C.tp]. Nous avons également fourni des résultats pour les transmissions à fibre faiblement multimode (gain d'un facteur 6-10) en traitant un point critique : les couplages modaux aux connexions et leur impact sur la qualité des systèmes [14.J.tp].

Pour le second point nous avons publié des résultats pour l'optimisation des systèmes utilisant une amplification optique à base de semiconducteurs (gain d'un facteur 3) et pour la compréhension de l'impact des effets non-linéaires de ce type d'amplification [8.J.tp]. Ces travaux s'inscrivent dans le cadre des projets Celtic+ : SENDATE, ANR : STRADE, ULTRAWIDE, FUI : 100GRIA, CALIPSO.

- **Capteurs photoniques et Imagerie non-conventionnelle au service de la biométrie, de la santé et des sciences du vivant.** En complément des systèmes imageurs usuels qui exploitent des mesures d'intensité, les travaux menés ces dernières années par la communauté capteurs n'ont cessé de montrer l'apport des systèmes imageurs sensibles à la phase qui permettent d'accéder à des propriétés extrêmement ténues des tissus. Dans un tel contexte, nous avons conçu et réalisé deux systèmes pour deux applications distinctes.

La première concerne l'authentification biométrique (doigt, iris) hautement sécurisée au niveau du capteur. Le système imageur élaboré consiste en un imageur PS-OCT (Phase Sensitive - Optical Coherence Tomography). Pour en tirer pleinement profit, nous avons proposé une méthode de fusion d'images en intensité et phase (brevet FR3041423). Les résultats obtenus démontrent des performances d'authentification basées sur l'empreinte interne équivalentes à celles obtenues sur l'empreinte externe par des capteurs standards ; la phase est ici exploitée pour pénétrer plus profondément sous la surface du tissu et donc l'étudier de façon plus globale.

La seconde application concerne la détection précoce et non invasive de pathologies. Pour cette application, l'introduction de nouvelles méthodes faibles coûts et hautement sensibles est l'enjeu majeur. Pour cela, nous nous basons sur l'approche de microscopie par ptychographie de Fourier récemment mise au point par des équipes de Caltech et Berkeley (2013), et avons réalisé un premier prototype produisant des images adaptées aux pathologies visées (malaria, leucémie...)

- **Sources lasers III-V : propriétés d'émission en présence d'un feed-back optique externe.** Les lasers à semi-conducteurs III-V de haute performance sont des composants de base indispensables dans les futurs réseaux hauts-débits de télécommunications par fibre optique, dont la conception et l'implantation nécessitent une meilleure compréhension de leur fonctionnement. Dans un système de transmission optique, le module laser est particulièrement sensible aux réflexions parasites issues de divers éléments optiques constituant le système. Des effets physiques complexes induits par ces réflexions (ROE : Rétroaction Optique Externe) peuvent considérablement modifier le comportement du laser. Notre objectif est de mettre au point un outil de simulation robuste et fiable permettant de compléter quelques travaux théoriques déjà réalisés sur le phénomène de ROE dans un laser Fabry-Pérot à mono-section en régime stationnaire [21.C.tp, 17.C.tp]. Nous avons élaboré, en généralisant un modèle classique dit à ondes progressives, un modèle applicable à différents types de lasers (Fabry-Pérot ou Distributed Feed-Back) avec structures longitudinales prédéfinies (mono- ou multi-sections actives) [7.J.tp].

## Simulations, propagation et navigation

(C. Letrou, N. Samama, T. Taillandier-Loize (depuis 2017), A. Vervisch-Picois).

Les ondes radiofréquences sont utilisées dans des contextes d'applications de plus en plus variés, soumis à des contraintes d'utilisation nouvelles (zones interdites pour les drones), à des exigences de fonctionnement en milieux couverts, à des besoins de détection de cibles Radar en environnement non dégagé. Nos travaux dans le domaine de la simulation électromagnétique d'une part, sur les techniques de positionnement et de contrôle de trajectoires d'autre part, visent à répondre à ces besoins.

- **Algorithmes rapides pour l'électromagnétisme.** La méthode de Lancer de Faisceaux Gaussiens (LFG) à partir de frames de Gabor, développée à CITI, a fait l'objet de développements nouveaux : - d'une part dans le domaine des antennes, dans la continuité de la thèse d'I.F. Arias Lopez (soutenue en juin 2013). La méthode de partitionnement spectral associée au LFG a été validée par la reconstruction de diagrammes de

rayonnement expérimentaux, et la possibilité de l'utiliser pour des transformations champ lointain-champ en zone de Fresnel a été illustrée [20.C.tp]. - d'autre part dans le domaine de la propagation en milieu urbain ou suburbain, pour des applications de type Radar terrestre, dans le cadre du projet ASTRID PECUMIA, en collaboration avec Thalès et la PME NOVELTIS. La méthode de décomposition adaptative a été développée et associée à un nouveau concept de "frame image" pour tenir compte de la présence du sol. Les résultats obtenus ont été validés par comparaison avec une méthode des moments et une méthode d'équation parabolique [14.C.tp].

- **Positionnement en milieux contraints.** Les systèmes de navigation par satellites présentent de très fortes limitations dans les milieux couverts dans lesquels les signaux restent trop faibles pour être détectés. De nombreuses communautés s'impliquent ainsi dans la recherche d'une solution dont l'universalité devrait être équivalente à celle de ces systèmes en extérieur. Nos travaux dans ce contexte reposent sur deux approches : le positionnement relatif d'un ensemble de noeuds communicants afin de produire un système sans aucune infrastructure ajoutée [18.C.tp]; il s'agit d'étudier la déformation géométrique du réseau par des mesures point à point de variation de distances. La seconde approche [10.C.tp] envisage une continuité entre extérieur et intérieur par l'intermédiaire de transmetteurs intérieurs de signaux compatibles avec les récepteurs GNSS actuels. Dans ce dernier cas, la différence de phases de deux signaux émis par une antenne double permet de définir des hyperboloïdes de présence, et l'intersection de plusieurs telles nappes (quadriques) donne la position. Le positionnement par mesure de différences de phases peut ainsi être assuré en mode dynamique avec une précision relativement bonne, les mesures de phases étant moins sujettes aux déformations que les mesures classiques [9.C.tp], comme les trajets indirects et la diffraction.
- **Navigation des drones.** La détection de drones survolant des zones interdites pose des problèmes scientifiques forts et des recherches sont menées afin de générer des zones d'exclusions. Nos travaux ont porté plus spécifiquement sur les systèmes de navigation par satellites (GNSS) et les techniques potentielles de leurrage (spoofing en anglais, à distinguer du brouillage, "jamming") de ces derniers afin d'empêcher l'intrusion [2.C.tp]. Les drones utilisent en effet les récepteurs GNSS pour calculer leur position, dont ils se servent pour définir le cap à suivre lorsqu'ils sont en mode de vol automatique au cours duquel le drone suit un plan de vol prédéfini à partir de points de cheminement (waypoints). Il s'agit ici d'émettre de "faux" signaux GNSS détectables et traitables mais qui induisent de fausses positions : si le drone les utilise, il pense être là où il n'est pas et ne se comporte pas comme prévu. On peut même dans certains cas extrêmes "contrôler" la trajectoire du drone en maîtrisant la fausse position calculée par le récepteur. Dans le même temps et en préparation de nos travaux futurs, nous avons développé de nouvelles métriques d'estimation de la précision d'un positionnement GNSS [6.C.tp] afin de disposer de divers outils de détection d'un leurrage (défense anti-leurrage).

## 7.2.2 Faits marquants de l'équipe TIPIC

L'équipe TIPIC a changé de périmètre : départ du groupe "apprentissage statistique et reconnaissance des formes"; intégration du groupe "technologies avancées pour les communications" (2015); recrutement de D. Clark (2017), Y. Petetin (2015) et T. Taillandier-Loize (2017). Cette restructuration a donné lieu à une première collaboration entre F. Lehmann et Y. Frignac (algorithmes de communication sur support optique). Pour la période à venir, la montée des activités en apprentissage statistique devrait engendrer de nouveaux échanges entre membres de l'équipe; par ailleurs l'arrivée des physiciens devrait permettre une meilleure intégration dans l'équipe de l'activité algorithmes rapides pour l'électromagnétisme, et des collaborations pourraient émerger entre spécialistes de la navigation et du traitement d'antennes (voir projet de recherche ci-dessous).

Parution de la monographie de recherche [1.L.tp].

B.-E. Benkelfat a été nommé Senior Member de l'Optical Society of America en 2013; D. Clark a été nommé Senior member IEEE en 2017; Y. Gottesman a reçu le prix du business impact et le prix d'excellence d'innovation pour le projet européen ITEA IDEA4SWIFT; l'article [4.J.tp] a reçu le prix du meilleur article algérien de l'année 2017 en Computer Science et Engineering; Florian Maire, qui a effectué sa thèse à TIPIC sous la direction de R. Douc (2011-2014), a reçu le prix DGA de la meilleure thèse en novembre 2016.

### 7.3 Organisation et vie de l'équipe TIPIC

L'équipe TIPIC a organisé les séminaires suivants :

- W. Pieczynski. Chaînes de Markov Triplet (06/2013).
- R. Sabourin, Université de Montréal, Canada. META-DES : A dynamic ensemble selection framework using meta-learning (06/2015).
- J.B. Courbot, CRAL observatoire de Lyon. Les observations astronomiques issues de l'instrument MUSE et leurs traitements (02/2016).
- E. Monfrini. Étiquetage morphosyntaxique par chaînes de Markov (04/2016).
- Y. Petetin. Introduction au deep learning (06/2016).
- G. Huyet, Cork Institute of Technology and Tyndall National Institute, Cork, Irlande. Properties of frequency-swept sources (12/2016).
- J. Lember, université de Tartu, Estonie. On segmentation with hidden, pairwise and triplet Markov models (06/2017).
- Y. Shmaliy, université de Guanajuato, Mexique. Robustness of the Kalman,  $H_\infty$ , and UFIR Filters (08/2017).
- M. Gomes-Borges, Safran. A Risk-Based Sensor Management using Random Finite Sets and POMDP (11/2017).
- S. S. Mukherjee, AnyVision, Grande-Bretagne. Face recognition that actually works (11/2017).
- D. Cormack, Leonardo, Grande-Bretagne. Sensor calibration from opportunistic targets in a maritime surveillance application (11/2017).
- A. Narykov, université Heriot-Watt, Grande-Bretagne. Sensor management and threat assessment (11/2017).
- B-N Vo, université de Curtin, Australie. Model-Based Learning for Point Patterns (06/2018).

Du fait de la large couverture thématique de TIPIC (voir le SWOT ci-dessous), l'interaction entre membres de l'équipe se fait également volontiers sous forme de petits groupes de lecture focalisés sur des sujets d'intérêt commun (optimisation convexe, apprentissage statistique, RKHS...).

### 7.4 Analyse SWOT de l'équipe TIPIC

L'analyse de nos activités, de nos résultats et de notre positionnement conduit à l'auto évaluation suivante :

Points forts	<ul style="list-style-type: none"> <li>— L'activité de TIPIC est équilibrée entre publications, valorisation et transfert, appui à la communauté, et formation par la recherche ;</li> <li>— plusieurs membres de TIPIC sont impliqués dans la vie scientifique internationale (membres de comités techniques IEEE, éditeurs associés de revues de référence : AAP, IEEE TSP, Signal Processing) ou nationale (ANR, GDR ISIS) ;</li> <li>— le positionnement de plusieurs membres de l'équipe TIPIC se situe à mi chemin entre théorie et pratique, ou entre théorie et expérimentation ; les aspects théoriques permettent ainsi de pérenniser les thématiques proches des applications, même si celles-ci évoluent ; les aspects pratiques sont souvent une source puissante d'inspiration ;</li> <li>— liens directs entre certains enseignements de M2 (Voie d'Approfondissement "Modélisations statistiques et Applications", Telecom SudParis, 3<sup>ème</sup> année ; Master ROSP de l'Université Paris Saclay) et certaines des thématiques de recherche de l'équipe.</li> </ul>
Points faibles	<ul style="list-style-type: none"> <li>— La relative diversité des thématiques ne favorise pas l'interaction forte entre les membres de l'équipe ;</li> <li>— difficultés à renforcer l'équipe avec des chercheurs CNRS.</li> </ul>
Opportunités	<ul style="list-style-type: none"> <li>— L'intégration dans la faculté "Data Sciences and Information Technologies" de l'institut NewUni, ainsi que le futur déménagement partiel du laboratoire vers le plateau de Saclay (2019), aideront l'équipe TIPIC, d'une part à se rapprocher de partenaires académiques privilégiés (LTCI, CEMAP, etc.), d'autre part à collaborer avec des acteurs industriels du plateau de Saclay.</li> </ul>
Risques	<ul style="list-style-type: none"> <li>— Compétitivité accrue des sources de financement nationaux et européens.</li> </ul>

## 7.5 Projet scientifique de l'équipe TIPIC

Le projet prend en compte les interactions potentielles entre membres de TIPIC dues au nouveau périmètre de l'équipe, au recrutement de 3 EC (voir la section "faits marquants"), et au souhait d'une certaine évolution des thématiques de recherche. Il est structuré en 3 axes ; en particulier les axes "Méthodes d'estimation et de détection pour les traitements multicateurs" et "Simulations, propagation et navigation" évoluent en un nouvel axe, consacré à la détection, l'estimation, la communication et la surveillance en environnements complexes. Ce projet devrait être renforcé par des recrutements d'EC : TSP a embauché en 2018 deux Maîtres de Conférences (en optique et en Internet des Objets) et un Professeur (en Apprentissage Statistique et Intelligence artificielle), qui ont tous trois vocation à rejoindre prochainement l'équipe TIPIC. Sur le plan thématique, le projet intègre la prolongation d'activités de recherche en cours mais comporte également une prise de risque sur des sujets nouveaux.

### Statistique et optimisation

(M. Castella, D. Clark, F. Desbouvries, R. Douc, E. Monfrini, Y. Petetin, W. Pieczynski).

Les membres de TIPIC spécialisés en statistique computationnelle et en optimisation ont accumulé depuis plusieurs années une expertise leur permettant désormais d'aborder des problèmes ambitieux, en particulier liés à l'inférence en grande dimension. Ils souhaitent par ailleurs profiter de leur maîtrise des modèles probabilistes markoviens pour établir des liens entre certains de ces modèles graphiques et des structures neuronales classiquement utilisées en apprentissage statistique et en intelligence artificielle.

#### — Simulation et optimisation en grande dimension.

Dans de nombreuses applications (données astronomiques en imagerie hyper spectrale, flux vidéo de caméras de vidéosurveillance, grands réseaux de capteurs, météorologie), l'état à estimer ou la fonction à optimiser sont de grande voire très grande dimension. Or les techniques classiques de statistique computationnelle ou d'optimisation ne passent pas à l'échelle, ce qui constitue un enjeu méthodologique majeur de la communauté Traitement du signal pour les prochaines années (voir le document de renouvellement du GDR ISIS<sup>1</sup>, partie prospective, thème A) Plus précisément, nous allons explorer dans cette thématique les pistes suivantes :

- *Algorithmes MCMC en grande dimension.* Nous nous proposons de remplacer la dynamique markovienne des algorithmes MCMC par des processus Markoviens déterministes par morceaux de même distribution de probabilité cible, dont nous essaierons de contrer la possible multimodalité par des dynamiques d'exploration possiblement inhomogènes dans l'espace ;
- *Simulation pour l'optimisation en grande dimension.* La piste que nous souhaitons explorer consiste à adapter à l'optimisation stochastique des idées provenant des algorithmes "pseudo-marginaux", qui sont une classe d'algorithmes de type MCMC dans lesquels on substitue à une loi cible incalculable un estimateur sans biais de celle-ci ; l'objectif final étant de concevoir, dans un contexte de grande dimension, des algorithmes d'optimisation en ligne efficaces ;
- *Assimilation de données.* Dans un contexte bayésien, le calcul numérique d'intégrales comporte le plus souvent une étape d'assimilation de nouvelles données. Cependant les techniques existantes sont très coûteuses à mettre en oeuvre dans les cas sévères (et en particulier en grande dimension), aussi nous souhaitons proposer des solutions plus efficaces en explorant des pistes telles que le choix des densités d'importance, des mécanismes de pondération, ou du mécanisme de transport des tirages Monte Carlo.
- *Inférence sur données massives corrélées.* L'inférence statistique sur les données massives fournies par les imageurs modernes est particulièrement difficile. Notre projet dans ce domaine consiste à capitaliser sur notre connaissance de modèles dynamiques complexes tels que les modèles de Markov triplet ; l'efficacité de nos algorithmes devrait résulter d'une prise en compte efficace de la structure de corrélation du processus caché. Les applications envisagées sont l'extraction de données sur cube d'images microtomographiques dans le cadre des maladies cardiovasculaires, l'imagerie astronomique hyperspectrale ou la reconnaissance automatique de cibles.
- *Interactions entre objets dans une scène multi-objets dynamique.* Nous nous proposons de développer une nouvelle classe d'algorithmes d'estimation multi-objets basés sur les cumulants et cumulants factoriels ; par rapport à l'état de l'art, ces algorithmes devraient simplifier l'évaluation des risques de collisions entre objets d'une même scène, en particulier lorsque le nombre d'objets est très élevé.

1. <http://www.gdr-isis.fr/uploads/documents/Renouvellement-ISIS-2019-2023.pdf>

— *Optimisation non convexe en grande dimension.* La difficulté principale en optimisation non convexe consiste à trouver un minimum global et non pas local, comme le permettent la plupart des méthodes actuelles. Nous nous concentrerons sur les optimisations polynomiale, rationnelle, ou plus généralement semi-algébrique; dans ce cadre, important d'un point de vue applicatif, il existe des hiérarchies de relaxations convexes du problème de départ convergeant vers l'optimum global cherché, mais au prix d'une dimension très grande des problèmes obtenus, souvent de type programmation semi-définie. Nous essayerons de dépasser les limitations des solveurs actuels en nous focalisant sur les méthodes du premier ordre et proximales.

— **Modèles probabilistes graphiques et structures neuronales.**

Les techniques d'apprentissage automatique, et en particulier les réseaux de neurones, ont connu ces dernières années un regain d'intérêt extraordinaire. Cependant, malgré des performances qui améliorent l'état de l'art dans de nombreuses applications, en particulier en traitement d'images dans le cas des réseaux de neurones profonds convolutionnels ou en traitement automatique de la parole dans le cas des réseaux de neurones récurrents, ces modèles neuronaux ne sont pas totalement compris d'un point de vue théorique. D'une part, l'impact du nombre de neurones et du nombre de couches sur l'approximation produite par les architectures profondes reste floue; d'autre part, il reste à comprendre la nature de la dépendance temporelle des données traitées qui est modélisée par les architectures récurrentes de type Long Short Term Memory (LSTM) et Gated Recurrent Unit (GRU).

L'intérêt de notre équipe pour les réseaux de neurones part d'une double observation: d'une part, les opérations complexes effectuées par certaines architectures neuronales font intervenir des variables intermédiaires cachées; d'autre part ces transformations sont récursives (- itérations d'une couche à la suivante dans le cas de réseaux de neurones profonds, itérations séquentielles dans le cas de réseaux de neurones récurrents), ce qui traduit implicitement une Markovianité sous-jacente. En définitive les réseaux de neurones présentent une structure de dépendance entre variables aléatoires proche de celle existant dans certains modèles (chaînes et arbres de Markov cachés et certaines de leurs extensions "couple" et "triplet") dont l'équipe est familière. Nous aborderons donc l'étude théorique des architectures neuronales au travers des liens qu'elles présentent avec les modélisations probabilistes markoviennes. Cette fertilisation croisée devrait bénéficier aux deux champs d'étude: l'objectif à long terme consiste à englober ces deux modélisations dans un cadre commun dans le but d'une part de construire des modélisations markoviennes plus puissantes, et d'autre part de mieux comprendre le résultat fourni par les architectures neuronales récurrentes.

## Détection, estimation, communication et surveillance en environnements complexes

(D. Clark, J.-P. Delmas, F. Lehmann, C. Letrou, N. Samama, T. Taillandier-Loize, A. Vervisch-Picois).

Le développement de réseaux de capteurs interconnectés capables de fournir une surveillance continue de scènes de grande taille suscite un intérêt croissant et nécessite la collecte, la communication et la fusion de données à partir de capteurs hétérogènes situés sur différentes plates-formes. Le caractère massif des données produites par les capteurs modernes, et les interférences liées à la nécessité de communiquer dans des domaines électromagnétiques encombrés, rendent nécessaire le développement de nouvelles techniques de communication, détection, estimation et surveillance dans des réseaux dynamiques de capteurs qui soient fiables, robustes et fonctionnant à différentes échelles. Nous nous proposons dans ce cadre d'explorer les pistes suivantes.

— **Gestion automatique de capteurs pour la surveillance d'objets multiples dynamiques**

Dans de nombreux domaines (notamment celui de la défense) il est devenu courant d'utiliser des réseaux de capteurs hétérogènes (caméras, radars), dont on souhaiterait désormais qu'ils prennent des décisions de façon automatique, sur la base des seules données accessibles sur les capteurs. Dans ce contexte un problème d'intérêt est l'allocation automatique de capteurs. L'approche que nous envisageons consiste à proposer des statistiques utiles dans différentes régions de la zone étudiée, et, plus précisément, à identifier quelle information nouvelle on recueille lorsqu'on observe de nouvelles régions. En termes de théorie de l'information, le problème revient non plus à calculer l'information mutuelle moyenne entre un état nouvellement observé et un état courant, mais entre les variables occupant une région d'intérêt, et celles occupant une autre région d'intérêt. Les outils classiques de théorie de l'information ne sont pas adaptés à ce problème multi-objets, aussi convient-il d'étendre la théorie existante au cadre des processus ponctuels et de la statistique multi-objets. Cette étude devrait donner lieu à de nouveaux concepts tels que l'entropie d'une région ou l'information mutuelle d'une région sur différents instants, et devrait permettre de construire une stratégie

de positionnement optimal des capteurs dans des réseaux dynamiques.

— **Estimation, détection et décodage conjoints dans les réseaux de type Internet des Objets**

L'émergence rapide de nouveaux concepts comme la ville intelligente, les véhicules autonomes ou encore l'industrie du futur repose sur le déploiement de capteurs ou d'actuateurs avec une autonomie énergétique ainsi qu'une densité par  $\text{km}^2$  inégalées, ce qui suppose que des milliers d'objets soient en mesure de communiquer de façon fiable malgré des ressources limitées dans les domaines temporel, fréquentiel et spatial. Ce problème d'accès massif représente un formidable défi que les réseaux de communication actuels sont incapables de relever. En particulier, un problème ouvert est celui de l'accès multiple pour des objets à émissions sporadiques pour lesquelles les protocoles existants de type handshake sont inadaptés car trop peu efficaces en termes de latence et de probabilité d'établissement d'une communication. Nous envisageons dans ce contexte de proposer des méthodes de traitement statistique originales permettant conjointement de détecter l'activité des objets, d'estimer le canal et de détecter les données transmises.

— **Positionnement dynamique par mesures de phase différentielles temporelles.**

Les systèmes de positionnement basés sur des signaux radio voient leurs performances dégradées dans des environnements tels que les canyons urbains ou l'intérieur de bâtiments, du fait entre autres du type de mesures réalisées, fondé en général sur une séquence qui permet tout à la fois d'identifier l'émetteur et d'estimer la distance avec le récepteur. Les méthodes de type RTK, basées sur des mesures de phase de la porteuse, sont en revanche moins sensibles aux erreurs (échos, diffraction, etc.). Notre approche dans ce domaine consiste à étudier ces systèmes à mesure de phase dans le mode différentiel temporel (car le calcul dans ce cas ne nécessite pas l'assistance d'une station de base). Nous étudierons notamment les performances de notre système en milieu extérieur (satellites et signaux de type GNSS), et en milieu intérieur (émetteurs à antenne double).

— **Algorithmes rapides pour l'électromagnétisme**

Les besoins en matière de détection dans des environnements complexes, de maîtrise des perturbations électromagnétiques, donc de capacité de prédiction de la propagation des ondes radio dans des environnements urbains ou intra-bâtiments sont de plus en plus élevés. Nous nous proposons d'une part d'utiliser notre méthode de Lancer de Faisceaux Gaussiens basé sur les frames de Gabor (LFG-fG) pour effectuer des calculs de surface équivalente radar monostatique aux longueurs d'ondes centi- ou millimétriques ; et d'autre part de dépasser les limites de cette théorie (dans le cas des objets diffractants non réguliers ou de dimension électrique insuffisante), en envisageant l'hybridation de cette méthode avec le "LFG avec redécomposition", l'Optique Physique multiniveaux ou avec une Méthode des Moments.

— **Détection d'interférences radio sur objets mobiles**

La multiplication de "brouilleurs" radio, intentionnels ou pas, constitue un enjeu majeur des années à venir, en particulier dans le cadre de l'explosion des systèmes rayonnants (internet des objets, drones, voitures,...). Le cas des drones modernes présente un intérêt particulier car il regroupe diverses problématiques : récepteur dédié à la navigation, systèmes de communication vers le sol et vers d'autres drones (vol en escadrille). Les interférences, tant pour la navigation que pour le contrôle ou les communications, seront un sujet d'étude. Nos travaux se focaliseront en particulier sur des méthodes de traitement d'antennes ; il s'agira de détecter et d'estimer la direction d'arrivée d'un ou plusieurs signaux d'interférence en présence de constellations GNSS standards. Par ailleurs, nous aborderons le comportement des horloges récepteur en situation de brouillage ou de leurrage, ainsi que la fusion instantanée des données de variation de phase des signaux GNSS et de la centrale inertielle. Les effets de la propagation sur le comportement des systèmes pour lesquels nous utiliserons les algorithmes cités précédemment seront traités afin de permettre des simulations réalistes de l'environnement.

## Optique et photonique

(B.-E. Benkelfat, Y. Frignac, Y. Gottesman, C. Lepers, Q. Zou).

Les travaux en optique photonique s'articulent comme dans la période précédente autour de trois axes. Les études consacrées au transport de l'information exploreront un champ nouveau, puisqu'elles s'orienteront désormais sur l'apport potentiel des algorithmes d'apprentissage automatique aux communications sur fibre optique et, réciproquement, à l'implémentation sur composants photoniques de tâches d'apprentissage automatique ; l'axe consacré aux capteurs photoniques se focalisera sur la microscopie sans marqueur par ptychographie de Fourier ;

enfin les travaux sur les lasers se focaliseront sur la dynamique de lasers à semi-conducteurs en présence de perturbations externes.

— **Bénéfices croisés de l'apprentissage automatique et des systèmes photoniques**

Les systèmes (i.e., liaisons point à point) et réseaux optiques ont vocation à devenir de plus en plus adaptables (notamment au trafic de données), avec en particulier l'introduction du SDN (Software Defined Networking), des transmissions cohérentes, etc. Dans ce contexte, les algorithmes d'apprentissage automatique devraient permettre de prendre en compte les données d'exploitation relevées sur les noeuds du réseau afin notamment de rendre ces systèmes et réseaux plus agiles et d'améliorer globalement leurs performances. Plus précisément, nous projetons dans les prochaines années :

- d'une part, d'appliquer des algorithmes d'apprentissage pour permettre aux transpondeurs optoélectroniques d'identifier automatiquement les codes de modulation transmis, le débit, ainsi que les dégradations affectant la transmission sur fibre optique, afin d'obtenir une auto-adaptation des paramètres de traitement des signaux et in fine d'améliorer la performance d'une transmission point à point ;
- d'autre part, d'exploiter conjointement, au niveau du réseau, les données de la couche physique (qualité du canal de transmission, alarmes, informations de contrôle) et les besoins des services en termes de trafic afin d'améliorer globalement l'exploitation des réseaux optiques (routage adaptatif, anticipation de pannes, reconfiguration, moindre consommation d'énergie).

Si l'apprentissage automatique peut se révéler bénéfique aux transmissions sur réseaux optiques, nous souhaitons également développer des **systèmes photoniques neuro-morphiques** permettant de réaliser des tâches d'apprentissage automatique de manière plus rapide et moins consommatrice en énergie que les processeurs de calculs graphiques actuels. Pour ce faire il s'agira dans un premier temps d'optimiser numériquement et de tester expérimentalement des techniques pour le "reservoir computing" photonique à base de composants optoélectroniques employés communément dans les télécommunications optiques.

— **Microscopie sans marqueur pour le diagnostic médical**

Nos études vont essentiellement se focaliser sur la microscopie par ptychographie de Fourier pour lequel un prototype vient d'être achevé ; l'enjeu consiste désormais à contrôler les erreurs introduites lors de la reconstruction des images en phase. Sur le plan méthodologique il convient de mieux identifier l'origine des bruits, par une étude du modèle du microscope, et d'autre part de renforcer la qualité de la phase reconstruite (clé pour accéder à la morphologie 3D des structures) ; nous proposerons pour cela des traitements alternatifs des données, basés notamment sur l'introduction d'une redondance partielle des informations acquises par contrôle de l'excitation optique des échantillons. Sur un plan opérationnel, nous projetons enfin d'appliquer notre imageur à la biologie fondamentale (aide au suivi de la division cellulaire de cellules souches ; il s'agit d'un problème fondamental ouvert pour lequel il n'existe pas de solution d'imagerie à ce jour), dans le cadre d'un partenariat avec le Généthon ; et à des problèmes de biologie appliquée tels que la détection précoce et non invasive de pathologies (paludisme, leucémie), conjointement avec les équipes ACMES et ARMEDIA de SAMOVAR, Mines-ParisTech, et la société TRIBVN.

— **Dynamique de lasers à semi-conducteurs en présence de perturbations externes**

Nous prolongerons l'étude en cours sur les propriétés d'émission de lasers Fabry-Pérot à mono-section en présence d'un feed-back optique externe. Nous nous intéresserons plus particulièrement au comportement dynamique (bifurcations, effondrement de la cohérence, chaos, etc.) de ces structures lorsqu'elles sont soumises à des perturbations d'origines physiques différentes (feed-back optique externe, injection optique, feed-back mixte optique/électrique, etc.), et sur une étude comparative du modèle de laser que nous avons développé et du modèle classique dit de Lang-Kobayashi.