HCERES

Haut conseil de l'évaluation de la recherche
et de l'enseignement supérieur

Auto-évaluation
HCERES vague D

2012 – 2017

Laboratoire d'Informatique de Paris 6

UMR 7606 CNRS – UPMC

UPMC
SORBONNE UNIVERSITÉS

cnrs

Contents

HCERES
Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur

LIP6 UMR 7606 CNRS - UPMC

Vague D
Campagne d'évaluation 2017 – 2018
Unité de recherche
Dossier d'autoévaluation

Informations générales

Nom de l'unité : Laboratoire d'Informatique de Paris 6
Acronyme : LIP6

Champ de recherche de rattachement : Informatique

Nom du directeur pour le contrat en cours : Jean Claude Bajard, adjoint Franck Petit
Nom du directeur pour le contrat à venir : Fabrice Kordon

Type de demande :

Renouvellement à l'identique ■ Restructuration ☐ Création ex nihilo ☐

Établissements et organismes de rattachement :

Liste des établissements et organismes tutelles de l'unité de recherche **pour le contrat en cours** et **pour le prochain contrat** (tutelles).

Contrat en cours :	Prochain contrat :
- UPMC	- SU
- CNRS	- CNRS
- . . .	- . . .
- . . .	- . . .

Choix de l'évaluation interdisciplinaire de l'unité de recherche ou de l'équipe interne :

Oui ☐ Non ■

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'unité

Introduction

LIP6 was founded the first of January 1997, and the result of the merger of three laboratories: the LAFORIA, the MASI and a part of the LITP. It represents more than 90% of the Computer Scientists of SU-UPMC. We are located on the Jussieu Campus in three buildings of UPMC, and we have around $6000m^2$ at our disposal for hosting all the LIP6 members, the experiment platforms, and the servers of the laboratory.

Since January 2006, LIP6 is organized in departments which gathers the 22 teams by thematics (3 are INRIA projects). The laboratory was initially composed of 5 departments, the biggest one was Networks and Distributed Systems and was split in two departments in 2014, Networks and Systems, and, Complex Systems. In 2014, the MALIRE team was split in two: MLIA (Machine Learning and Information Access) and LFI (Learning Fuzzy and Intelligent systems). Whisper a new INRIA team was created in 2014 as a spin off of the INRIA team REGAL.

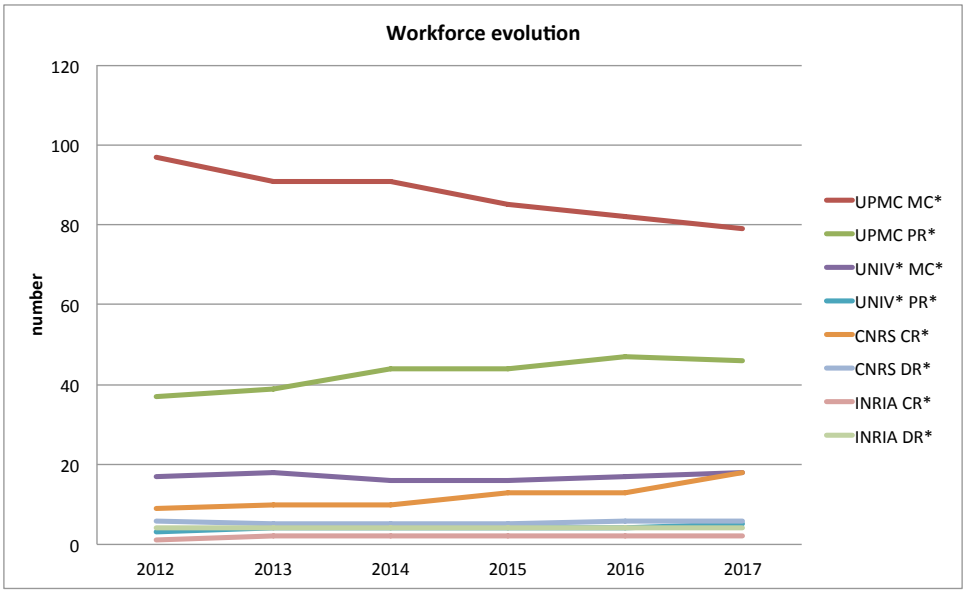
A new team ALMASTY in Networks and Systems dept. (NetSys) was created in 2015 and LIP6 welcomed the Quantum Information team coming from LTCI (lab of Telecom ParisTech). This team integrated the Systems On Chips dept. (SOC).

LIP6 offers to the researchers and to the teams "official" web pages and an access to the data bases of the laboratory. Also, all the departments offer different common services: email, servers accounts, web accounts... The laboratory owns some experimental platforms like GPU or CPU clusters (big and small), some teams provide all the laboratory with an access to their different servers. The computer service supports the researchers on different kind of applications: Network, Software, Embedded System, Hardware and the clusters.

The LIP6 hosts, since 2011, the FIT equipex (Plan Investissements d'Avenir) which is a testbed for Networks experimentations opened to the research community, Academic and/or Industrial, national and/or international. FIT is now an Infrastructure de Recherche (Ministère de l'Enseignement, de la Recherche et de l'Innovation).

Tableau des effectifs et moyens de l'unité

Human resources. Over the last five years, the number of **permanent academic researchers** (universities, CNRS, INRIA,...) is stable, around 179, including the two professor chair holders. During this period, ten researchers left temporary our laboratory, sometimes for a long time, four are in secondment (two of them from UPMC and two from CNRS) and six on leave (four of them from UPMC).



In fact, this stable number is mainly due to the arrival of 9 junior and one senior researchers from CNRS. The number of UPMC assistant professors, including secondment and leave of absence, decreased significantly

from 97 in 2012 to 80 in 2017. At the same time, the number of professors increased from 37 in 2012 to 46, these numbers include the two chairs offered by the *fondation partenariale de l'UPMC*. The number of professors and assistant professors from other universities was fairly stable, around 23 (5 professors and 18 assistant professors).

Organization	Type	2012	2013	2014	2015	2016	2017
UPMC	MC*	97	91	91	85	82	80
UPMC	PR*	37	39	44	44	47	46
UNIV*	MC*	17	18	16	16	17	18
UNIV*	PR*	3	4	4	4	4	5
CNRS	CR*	9	10	10	13	13	18
CNRS	DR*	6	5	5	5	6	6
INRIA	CR*	1	2	2	2	2	2
INRIA	DR*	4	4	4	4	4	4
Others		2	2	1	1	1	0
Total		176	175	177	174	176	179

Regarding the **PhD students**, we could note some fluctuations between 47 to 61 new students per year. The average duration of a PhD, from the beginning to the defense is three years and half, which is correct but perfectible. Almost all our PhD students defend. The 2017 values are not significant as the university year starts in September and most of the defenses are in the fall, thus they do not appear in the table.

PhD	2012	2013	2014	2015	2016
in progres	196	198	192	191	178
first year	49	61	58	47	48
defense	48	65	58	62	44
duration	3.89	3.66	3.27	3.68	3.42

Research is supported by an **administrative and financial service**, a **computer service** and a **manager** (IR UPMC). The 30th of June 2017, the administrative and financial service (SCA) is composed of 9 persons (7 UPMC, 2 CNRS), and 13 engineers for the computer service (SCI) (8 CNRS, 5 UPMC). We have one **graphic designer** who is in charge of the communication, and one **project manager** (engineer UPMC) who supports researchers for writing their proposals.

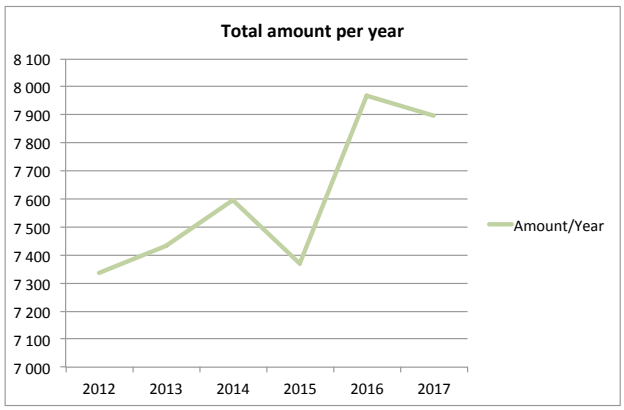
We develop a little bit more in the governance section of this report.

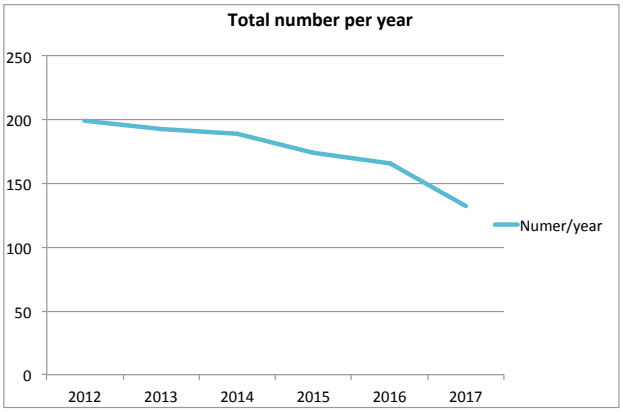
Organism	Grade	2012	2013	2014	2015	2016	2017
CNRS	IR	1	1	1	1	2	2
	IE	7	7	7	7	7	6
	AI	0	1	1	1	0	1
	TCH	3	3	3	3	2	2
CNRS	total	11	12	12	12	11	11
UPMC	IR	2	2	2	3	3	4
	IE	2	2	2	2	2	2
	AI	2	2	1	1	0	0
	TCH	4	5	5	5	6	6
	AJT	2	2	1	1	0	0
	AAEN	1	1	1	1	0	0
	SAEN	1	1	1	1	1	1
UPMC	total	14	15	13	14	12	13
LIP6	Total	25	27	25	26	23	24

The 30th of June 2017, LIP6 hosts within the research projects 37 **post-doctoral researchers** and 32 **engineers** (CDD). Those numbers change during the year depending of the project needs, but these values are representative. We hosted along the year training master students with a peak on spring around 106 persons for 2017.

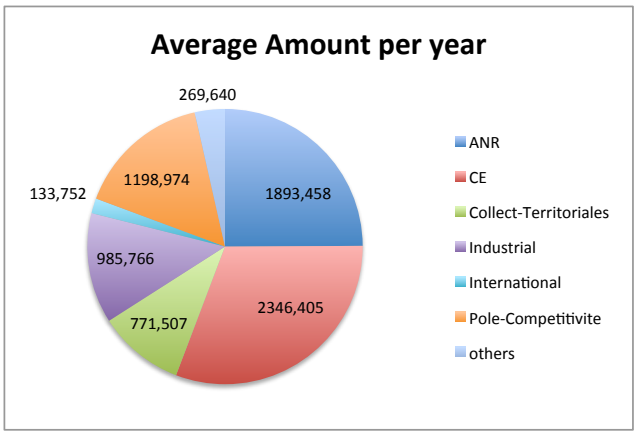
Financial supports. LIP6 belongs and manages a large number of National, European, and International collaborative projects. Most of them are including industrial partners but sometimes, they are purely industrial in nature. This ensures to the research teams an amount of annual resources of more than $7,5$M euros. This annual amount is quite stable, despite a slight decrease of the number of active contracts over the years.

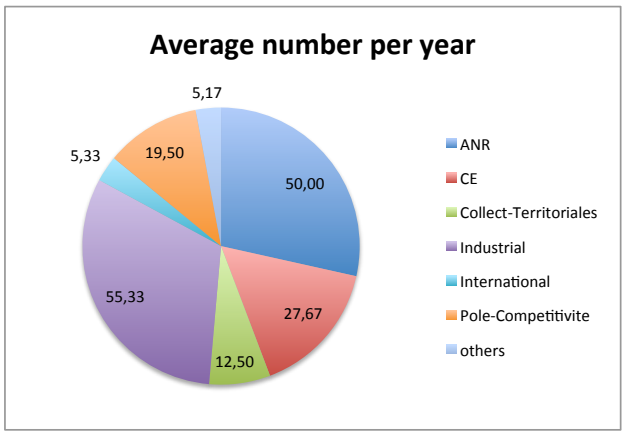
All the LIP6 grants are managed by UPMC, due to an agreement between CNRS and UPMC. Since 2017, a similar agreement had been signed with INRIA.



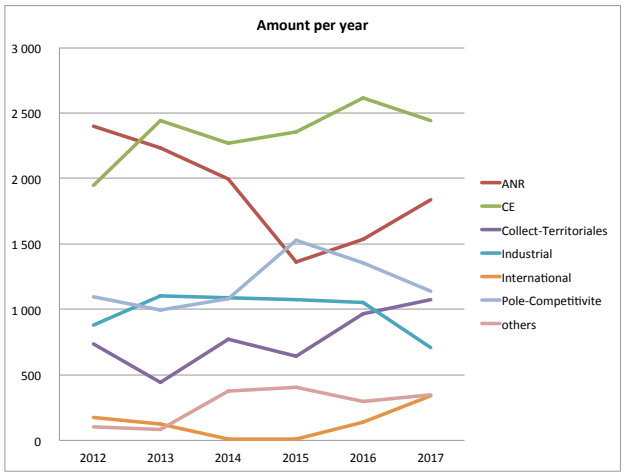


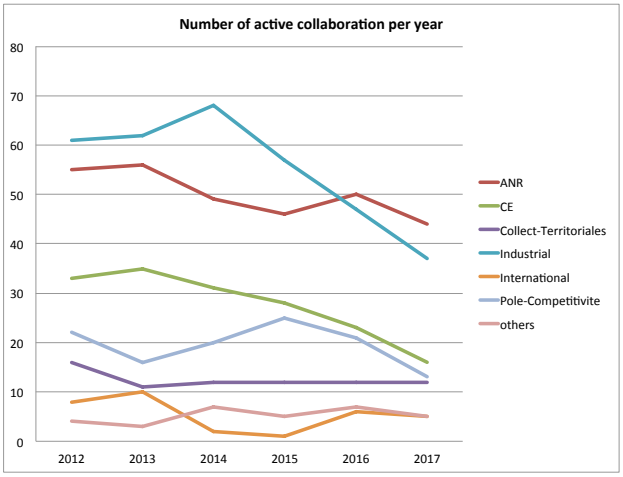
We notice that more of 56% of our financial supports come form the French national agency (called ANR) and the European Commision even if they represent 44% of the number of grants. Conversely, industrial collaborations which represent 38% in terms of number, are only 13% of the total financial amount. This is mainly due to the PhD industrial support grants (CIFRE) as their research support is around 15keuros each year.





We note a general decrease of the grant number in each category, but the amounts remain similar, even increase in some cases. This is due to an improved collaborative work with the DGRTT (Direction Générale de la Recherche et du Transfert Technologique).





The UPMC financial support is relatively stable since 2014 (in 2017 we received a first payment of 80%, the rest is in general given later). The CNRS allocation increases regularly to reach and pass the 50% of the UPMC one.

Organization	2012	2013	2014	2015	2016	2017
UPMC	360 284	277 933	359 000	359000	345 040	349 000
CNRS	129 000	136 000	145 000	150 000	160 000	180 000
UPMC platform grants	44 466		15 000			
CNRS platform grants		20 000	20 000	20 000		15 000

These allocations (UPMC and CNRS) enables the laboratory to take in charge common operational expenses and to support the research policy by allocating budgets for specific actions. Most significative numbers are given in the following table.

Actions	2012	2013	2014	2015	2016	2017
LIP6 projects	85 000	60 000	60 000	60 000	60 000	60 000
Soutien à la publication			6 000	8 206	12 000	22 000
Colloquium + events			17 000	20 000	15 000	15 000
Talent supports + newcomers			12 319	20 000	20 000	20 000
PhD student seminars					5 000	5 000
Invited Prof.						20 000
Team supports	350 784	253 433	306 500	249 000	215 040	217 000

Then, in 2014, the directors proposed to initiate a tax of 5% on the indirect costs of European projects and industrial projects, which was approved by the laboratory council (conseil de laboratoire). This measure was effective in 2015, and we received 30keuros in 2016 and 48keuros in 2017. This ressource helps financing some salaries for support personnel as our project manager. This employee contributes a lot to our incitative policy for submitting applications to European calls with a focus on ERCs, but also bilateral industrial collaborations.

Politique scientifique

Warning: we only cite some examples, more cooperations, events and details can be found in the team chapters.

The LIP6 is a joint unit research between Pierre et Marie Curie University and French National Center for Scientific Research (CNRS), UMR 7606 UPMC - CNRS, is a computer science research laboratory dedicated to the modeling and the resolution of fundamental problems driven by applications, as well as to the implementation and the validation through academic and industrial partnerships.

The main focuses are :

- Safety, security and reliability;
- Data science, intelligence and optimization;
- Smart devices.

LIP6 addresses these challenges at different thematic levels within its 6 departments and its 22 teams (3 are INRIA projects).

Scientific policy. It was initiated and driven by the directors and the **scientific council** of the laboratory. This council is composed of the 22 team leaders and the management of the laboratory. The head of the master degree with the head of the doctoral school are permanently invited as well as the departments' managers.

The scientific council particularly ensures the monitoring of the team evolution. In particular, it takes care of the recruitment with respect to the main laboratory trends. Also, it manages the selection of LIP6 project as well as the LIP6 invitations. It is a fundamental body for maintaining a coherent scientific policy.

The LIP6 management launched, since its arrival, a **publication support committee**. With a dedicated budget, it supports research activities which are temporarily not funded, as well as thematic transitions which are of particular interest for the laboratory. The outcomes are very positive. Researchers do not hesitate to contact this committee which allows to push new researches. Before, there was a committee that was only dedicated to non-publishing researchers without any funds.

Also, we promoted an incitative policy for helping to set up and submit applications to European calls and industrial collaborations. We hired a **project manager**, whose mission is to monitor the different calls, to prospect companies, to organize meetings, and to provide applicant researchers with help. Her work highlights the multiple opportunities for the researchers to discover all the possible cooperations, as well as the financial tools, but also to prepare their proposals and to improve relationships with the DGRTT (direction générale de la recherche et du transfert technologique) which finalizes and completes the submissions. Since her arrival in February 2016, many researchers who had never submitted to European calls, participated. First successes are coming: this position is a good operation for the laboratory.

With our project manager, we push our researchers to apply to ERC. Regarding this, we also created a small committee composed of Marc Shapiro, the director, the project manager, and successful ERC candidates. The mission of this committee is to prospect inside the laboratory who can be good candidates, to invite researchers to present their projects, to review proposals, to train for interviews. Moreover, we initiate financial supports to the candidates who can attend some meetings in Brussel and use the help of an external consultant.

Furthermore, in April 2014, the directors suggested to the management council, after a letter of UPMC President indicating a potential audit of the computer science at UPMC, to have our own **Scientific Advisory Board** (SAB). That was approved in September 2014 by scientific council of LIP6. Then the President of UPMC proposed to replace the audit by a SAB. At the end, the SAB that we tried to constitute became the one of the computer science at UPMC, knowing that LIP6 represents around 90% of the computer science activities. This SAB took place in October 2016 and its conclusions are more recommendations than evaluation. They underline that the department organization is not so fine, there is a risk of "balkanization". They suggested to suppress this level, and to give more importance to the scientific axes. This idea is now being discussed and could be applied by the future head of LIP6. The SAB was composed of Serge Abiteboul (chair, member of INRIA), Rina Dechter (UC Irvine, USA), Jan Hesthaven (EPFL), Anne-Marie Kermarrec (IRISA, INRIA), Muriel Medard (MIT, USA), Jean Walrand (UC Berkeley, USA).

All our collaborative agreements are supervised by the DGRTT (*Direction générale de la recherche et du transfert de technologie*). DGRTT takes care of the researchers and of the laboratory property rights, as well as University ones. They assist the researchers for the establishment of invention writings, of patent registrations. They help researchers to prepare creation of start-up. Then, SATT Lutech, a private body which works with universities, supports the researcher until the effective creation of a new entity.

Between 2012 and 2017, 29 softwares have been produced by LIP6 researchers and 17 patents have been registered. We have 9 filings of invention during this period, and, 4 start-ups have been created.

2. Produits de la recherche et activités de recherche

Bilan scientifique

LIP6 is a major actor of the computer science research at the regional and national level, as well as European and international ones. In 2017, more than 22% of the ANR and 44% of the European projects, in which LIP6 is involved, are managed by our laboratory. That shows the dynamism of our teams to establish collaborative projects and to lead National and European research initiatives. Our three mains axes of research are found in almost all the LIP6 projects (about fifteen per year) as well as all the collaborative projects where LIP6 is acting. It is clear that we should expect more events around these thematics launched in 2015, but such initiatives are taking time and our researchers are really busy, in particular assistant professors and professors with the teaching. But these three axes are now well established among researchers, and we are optimistic about the development of them.

The directors initiated and were involved in the preparation of several submissions like DIM *RFSI* (Domaine d'Intérêt Majeur, Réseau Francilien en Sciences Informatiques), Tremplin Carnot *Interface*, IUIS (Institut Universitaire de l'Ingénierie de la Santé), etc.—please refer below for more details. Most of the LIP6 teams are involved in these projects. Below are also provided some examples of the implication of LIP6 in the enhancement of research networks. This list is not exhaustive but gives a good idea of the LIP6 activities.

Inside Sorbonne Université (SU). the LIP6 is deeply implied in several SU initiatives: LabEx, EquipEx, SU Institutes, Carnot label... Several teams from electronics to artificial intelligence through distributed algorithmic, are involved in **LabEx SMART** driven by Prof. R. Chatila (ISIR). Many projects between LIP6 teams

and other SU laboratories, offered fruitful collaborations with interesting results. Inside **LabEx CalSimLab**, LIP6 collaborates with theoretical chemical lab., applied math laboratories, observatoire de Paris, ... LIP6 is in the leading committee of these two LabEx. Moreover, LIP6 is represented in the LabEx ObVil by Prof. J.-G. Ganascia who is involved in numerical humanities and established important links with the Paris-Sorbonne university (Paris 4).

The **EquipEx FIT** coordinated by Prof. S. Fdida, is an experimental network testbed involving many international collaborations. This testbed enters in a larger sharing international consortium named OneLab hosted at LIP6.

Collaborations built inside the LabEx CalSimLab take advantages of the **EquipEx equip@meso** and of all the resources of the **ISCD** (Institut des Sciences du Calcul et des Données), created by Prof. P. Frey (LJLL). Pequan team is one of the most active partner of the ISCD, with for example, the visualisation works of J. Tierny, using the image wall of this institute.

LIP6 is a partner of the **IUIS** (institut universitaire de l'ingénierie de la santé) and member of its leading committee, where many collaborative projects are initiated, involving in particular the teams SYEL, ACASA, MLIA...

Then, LIP6 belongs to the **label Carnot "émergent" Interface**, and is one of the leading actor with respect to its industrial contracts. Prof. A. El Fallah Seghrouchni is deputy head of this structure driven by Prof. G. Morelle (ISIR).

Regional Actor. LIP6 is involved in two main competitive **poles of the Ile-de-France region: Cap Digital and Systematic**.

The mission of Cap Digital is to launch some initiative around Parisian area industries of the numerical inventions. Professor J.M. Labat represents the UPMC in the administrative committee of this structure and professor S. Fdida is member of the executive committee.

Systematic is dedicated to complex systems, software and digital technologies. Professor E. Chailloux is member of the leading committee of the thematic group "Free Software".

LIP6 participates in several consortia applying to **DIM (Domaine d'Intérêt Majeur)** of the Ile-de-France region, three of which succeeded. DIM "*émergent*" **RFSI** (*réseau francilien en sciences informatiques*), led by Olivier Cappé, which gathers the computer sciences laboratories of Ile-de-France as well as IT clusters such as Cap Digital and Systematic, the IRT (*instituts de recherche technologique*) SystemX and the three SATT (*Sociétés d'Accélération du Transfert de Technologies*) Lutech, IdF Innov and Paris-Saclay. Thematics fit the main LIP6 axes. The Quantum Information team is deeply involved in DIM **SIRTEQ** (*Science et Ingénierie en Région Ile-de-France pour les Technologies Quantiques*), and, the ACASA team belongs to the DIM "*émergent*" **PRADA** devoted to text science ans new knowledges.

LIP6 is involved in several groups of interest in Paris. QI team is member of **Paris Centre for Quantum Computing** (PCQC) , where Eleni Diamanti is deputy. It will be hosted at UPMC nearby the LIP6. IRILL (*Initiative de Recherche et Innovation sur le Logiciel Libre*) hosted at UPMC; involves the teams APR and Whisper, APR is one of the main contributors. Then the **Laboratory of Information, Networking and Communication Sciences** (LINCS), located place d'Italie, implies most of the NPA team members. Pr. S. Fdida is member of its Executive Committee.

National Actor. Many members of the LIP6 are requested for their expertise by **national bodies** such as ANRT, ANR, AERES then HCERES, CNRS (Co-cnrs section 6, INS2I, PEPS...) Some participated to scientific committees of these agencies, for example with the ANR: CES38 2015, CONTINT 2013, VP du CE 38 2017,...

Regarding CNRS, several members are leaders of poles or working groups in different **GDR (*Groupement de Recherche*)**:

IM *Informatique-Mathématique* (Michèle Soria was co-head of Aléa),

GPL *Génie de la Programmation et du Logiciel* (P. Poizat drove COSMAL, J. Malenfant co-head of ALROB),

IGRV *Informatique Géométrique et Graphique, Réalité Virtuelle et Visualisation* (J. Tierny co-head of the working group VISU),

RO *Recherche Opérationnelle* (P. Perny co-head of the pole DMPE, P. Fouilhoux and E. Hyon are part of the management committee, animation of GT POC, P2LS, COSMOS and AGAPE),

IA *Intelligence Artificielle* (P. Perny co-head of the pole TADJ, N. Maudet member of the animation and scien-

tific committee, F. Pachet member of the scientific committee),
RSD *Réseaux et Systèmes Distribués* (S. Secci belong to the executive committee comité and is co-head of the pole Virtualisation, C. Magnien belongs to the executive committee of the pole ResCom, was member of the scientific committee of ASR : S. Fdida, S. Tixeuil, M. Dias de Amorim),
SoC/SiP *Logiciel embarqué dans des SOC/SIP* (B. Granado belongs to the executive committee, M.-M. Lourérat head of GT Système hétérogènes),
ISIS *Information, Signal, Image et ViSion* (M. Cord co-head of IRIM),
Madics *Masses de Données, Informations et Connaissances en Sciences* (B. Amann belongs to the heading committee),
IQFA *Ingénierie Quantique, des Aspects Fondamentaux aux Applications* (E. Diamanti belongs to the executive committee)...

Furthermore, LIP6 is involved in *CLEAR*, a common laboratory created in partnership with *Thales Group*. CLEAR mainly involves researcher from MLIA, LIA, and CN teams. It enables to carry out work on the perception, analysis, and control of information. These topics are dealt within various aspects of machine learning, artificial intelligence and networks.

International Actor. Regarding the European projects involving the LIP6, we currently have 16 in 2017 and we had a peak in 2013 with 35. The LIP6 european collaborations cover almost all the European countries.

Added to these European collaborations, a lot of international cooperations are established through different kinds of funding with PICS CNRS, GDRI CNRS, or Campus France (Projets Hubert Curien...), Capes-Cofecub, and the international ANR project like PRCI . LIP6 covers the different continents.

For example:

- North America: Chicago USA (France-US Partner University Fund, Fondation Maison des Sciences de l'Homme, fondation Mellon et la FMSH...), Washington USA (ANR MRSEI...), Canada (PICS FERRARI, ANR international France-Canada...),
- South America: Brasil (CNRS WINDS, Capes-Cofecub, CNRS-Inria-FAP's, SU - FAPERJ,Sciences sans Frontière ...), Chili (équipe associée INRIA...), Mexico (ANR International France-Mexique...),
- Asia: Israel (PHC Maimonide...), Japan (Tohoku , ARNI, ANR-10-JST-003...),Singapore (ANR PRCI), Viet Nam (PICS...), India (IFCPAR/CEFIPRA...), Iran (PHC-Gundishapour...), China (FP7 IRSES...), Corea (Illumine...), Tawain (ANR-MOST...),
- Africa: Tunisia (PHC-Utique,...), Algeria (CMEP Tassili 11MDU837,...),
- Oceania: Australia (PICS, FP7 NoE...),

We have a lot of invited professors or post-doctoral researchers coming from many different countries (123 Frenchs, 144 Europeans, 341 non European): Algeria, Australia, Austria, Brazil, Canada, Chile, China, Czech Republic. Danemark, Germany, Greece, India, Iran, Ireland, Island, Italy, Japan, Kenia, Luxembourg, Maroc, Netherlands, New Zeland, Norway, Poland, Portugal, Rumania, Russia, Senegal, Singapore, Sweden, Thailand, Turkey, Tunisia, UK, Uruguay, USA, Venezuela, Vietnam, ...

Données chiffrées

Publications - scientific production 2012-2017

The complete listing of the publications are one https://www.lip6.fr/production/publications.php

Type	2012	2013	2014	2015	2016	2017	Total
Jounals	134	122	121	131	159	60	727
Conferences	336	386	350	326	348	129	1875
Posters	6	9	11	6	11	2	45
Books and Chapters	26	26	20	18	22	13	125
Edition	8	8	11	9	11	4	51
PhD	47	68	58	62	44	9	288
HDR	9	5	4	5	9	2	34

The ratio conferences/journals is around 2.5 in average, with a production quite stable from one year to another. The average number of conferences per permanent researcher is higher than 10 (around 2 per year) and for the journal around 4 (around 0.8 per year). Of course, PhD students and post-doctoral researchers participate a lot to this production. As their number is almost equal to the number of permanent researchers, we can conclude that we have in average around 5 conferences and 2 journals per LIP6 member.

Each team presents its own contributions and pointed out the most significant activities in the report.

Grants. Here, we give the number of grants obtained each year by the LIP6. For this analysis, we only took into account projects that already started, therefore the data regarding 2017 are not complete yet. The annual numbers are regular, for each category the variation can be explained by the duration of the project (a team does not necessarily need new contracts each year) but also by some variations of the selection. We remind that the smooth annual total funds stay stable around 7.5Meuros each year.

Starting	2012	2013	2014	2015	2016	2017	Total
ANR	9	14	8	13	14	3	61
CE	7	5	2	7	8	4	33
Contrats de Recherche privés	19	13	20	12	14	2	80
International	4	5	0	1	5	1	16
Collect-Territoriales	0	3	5	4	1	1	14
Pole-Compétitivité	6	3	7	7	3	1	27
Autre	1	0	5	1	3	2	12
Total	46	43	47	45	48	14	243

Sélection des produits et des activités de recherche

This part is developed by each team of the laboratory, we just remind here the number of patent registrations and software production which are some of our strengths.

Type	2012	2013	2014	2015	2016	2017	Total
Patent	2	1	1	1	7	5	17
Sofware	7	3	3	7	5	4	29
Coproperty	1	0	0	0	1	0	2
Exclusive	0	0	6	0	0	0	6
Transfer of rights	0	1	0	0	0	0	1

We can note that we have 9 fillings of invention which are running, some could generate to a patent and/or a start-up. Please also note that 4 start-ups have been created, as it is shown in more details in the next section.

Faits marquants

Although it is challenging to summarize the remarkable facts of a big laboratory like the LIP6, we can here mention a few examples:

National and European Individual Grants

- **ERC - European Research Council Grants**
  - **François Pachet** (August 2012, Project ID: 291156): ERC **Advanced**, FlowMachines. Flow machines are interactive systems that learn how to generate content, text or music, in the user's style. http://cordis.europa.eu/project/rcn/104383_en.html
    FlowMachines was selected for a proof of concept year (March 2015 Project ID: 641187) http://cordis.europa.eu/project/rcn/193789_en.html
  - **Antoine Joux** (January 2016, Project ID: 669891): ERC **Advanced**, Algorithmic and Mathematical Cryptology (AlmaCrypt). Today's internet is protected by a form of cryptography based on complexity theoretic hardness assumptions. The main ambition of Almacrypt is to solve this

issue by challenging the assumptions through an advanced algorithmic analysis http://cordis.europa.eu/project/rcn/199377_en.html

  - **Antoine Miné** (June 2016, Project ID: 681393): ERC **Consolidator**, Modular Open Platform for Static Analysis (Mopsa). The Mopsa project aims at creating methods and tools to make computer software more reliable. http://cordis.europa.eu/project/rcn/202577_en.html
  - **Eleni Diamanti** : ERC **Starting**, 2017, QUSCO - Quantum superiority with coherent states.

- **LIP6 hosts three chairs (professorship)**
  - **Giovanni Pau**: UPMC, ATOS and Renault Chair: Mobility and Connectivity challenges (*La voiture connectée*)
  - **Antoine Joux:** (prix Gödel 2013, IACR Fellow 2014): chair Fondation partenariale UPMC : Cryptology
  - **Vanda Luengo**: chair *Environnements Informatiques pour l'Apprentissage Humain* (Computer environment for human learning).
- **Seven IUF since 2009** (M. Cord, G. Pujolle, M. Safey El Din, S. Tixeuil, J.G. Ganascia, A. Miné, D. Vergnaud)

National Programs of Excellence (PIA)

- **Plateforms - Testbeds**
  - **Future Internet of Things** (coordinator) (February 2011). FIT is provided by a consortium of five institutions: UPMC, INRIA, Strasbourg Univ., Institute Mine-Telecom, CNRS. Serge Fdida, member of the team NPA of the LIP6 is the coordinator. https://fit-equipex.fr
  - Equip@méso, scientific computations (February 2011). LIP6 participate through collaborations with ISCD (Data and Computing Sciences) of UPMC. http://www.genci.fr/fr/content/equipmeso-0
- **Laboratories of Excellence**
  - SMART (March 2012): human/machine/human interactions in the digital society. The partners are ISIR (Robotics), LIP6 (Computer Science), L2E (Electronics and Electromagnetism), LJLL (Applied Math), STMS (Music and Sound, IRCAM), LIB (Biomedical Imaging) from UPMC, Chart-Lutin from Cité des sciences et de l'industrie, LTCI from Telecom parisTech. Teams of the LIP6, involve in projects: MLIA (ONBUL), LFI, SYEL (SeNSE, SpinalCOM),NPA (Smart-Ban). http://www.smart-labex.fr
  - CALSIMLAB (March 2012), Simulation for scientific computing. The partners are ISCD (Computing institute), LCT (Chemistry), LCQB (Bio Computing), LPTMC (Chemistry), LIP6 (Computer Science), LJLL (Applied Math), IJLRA (mechanics), LOMIC (marine biogeochemistry), ISTeP (thermo-mechanics) from UPMC. Team of the LIP6 involves: PEQUAN. http://iscd.upmc.fr/research/labex-calsimlab/
  - OBVIL (March 2012), Numerical humanities. The partners are: CELLF (French Literature), CRLC (literature), CRIMIC (Spanish culture), CLEA (Spanish literature), PRITEPS (Theater), VALE (English literature), ELCI (Italian culture) from Paris Sorbonne, LIP6 (Computer Science) from UPMC and Bibliothèque Nationale de France (BNF). Team of LIP6 involves: ACASA. http://obvil.paris-sorbonne.fr

Start-up. LIP6 is used to the creation of start-ups, in particular in the SOC domain. During the period 2012-2017, we had four new start-ups born in our laboratory:

- **Flexras created** in 2009, founded by Hayder Mrabet then sold in 2015 to Mentor Graphics
- **Intento Design** created in 2015 by Ramy Iskander
- **Seamless Waves** created in 2016 by Hassan Aboushadi
- **Wisebatt** created in 2016 by Wilfried Dron (KIC EIT digita)
- **Blue Communication** reated in 2016 by Guy Pujolle

- "**JRC SA**" (Just-Right Consistency) creation in progress with INRIA and IT Translation. Marc Shapiro (INRIA)

3. Organisation et vie de l'unité

Pilotage, animation, organisation de l'unité

Strategic councils - Conseils

Laboratory council –conseil de laboratoire– composed of 20 persons, two third of elected members and one third of appointed members, deals with laboratory budget, the functional rules,... Its composition and its function are defined in the laboratory status and were voted in general assembly in 2014. For its tasks, it is helped by the different committees and the scientific council which transmit their work results. The Laboratory council is asked by the director for advices regarding current affairs of the laboratory life. It idiscusses the creation of the intern committees and its missions and therefore defines their internal functioning.

The **scientific council** is the organ where all the main scientific orientations are discussed. It deals with the definition of the skill profile for new professor or assistant professor positions and the sorting of them. It is also involved in the scientific life of the laboratory, manages the invitation of foreign researchers by the laboratory, the applications to LIP6 project fundings. It is composed by all the team leaders and the direction of the laboratory. The head of the doctoral school EDITE and the department managers are invited.

Then, the **management council**, composed of the department managers and the direction, helps the direction for urgent demands, the preparation of some laboratory or scientific councils.

Internal policy

The aim of the current management team is to give to the laboratory **an identity and an unity**. We developed different support actions for all laboratory members.

As said in the scientific policy section, scientific council has defined in 2014 the research activities as "dedicated to the modeling and the resolution of fundamental problems driven by applications". This is the genuine specific feature of our entity where theory works together with the applications, and involved in. We identified three main research axes which represent a large part of our activities:
- Safety, security and reliability;
- Data science, intelligence and optimization;
- Smart devices.

Since 2014, **newcomers** arriving in our entity benefit each of a welcome grant of 2000 euros. We organize a "welcome day" where we present the laboratory and where each newcomer gives a talk to introduce his career, his/her scientific developments and his prospective researches. It is also the opportunity to meet all the members of the laboratory around a buffet and to introduce the newcomers in a festive moment.

In 2014, we launched a campaign to rethink the distribution of space provided by UPMC. We defined, in laboratory council, new rules of distribution of the offices to the teams, giving to each team a vital space in function of the number of permanent members. This helped to maintain at the laboratory level at least 20% of the space as buffer area. This buffer space is managed by a **logistics committee** which suggests on demand, complementary space, if a team does not have enough space to receive a new PhD, post-doctoral or anyone else. This policy ensures a better space management, it avoids congestion of some teams and empty space in some others. That is perfectible, thus the committee proposes some adjustments of the rules to the laboratory council which can redefine the rules if needed.

In order to allow each member to pursue his research activities despite the randomness of the success at the multiple grant calls, we created in 2014 a **support committee for publication** which has its own budget, voted in the laboratory council. This committee supports the conference fees (flight, hostels, subscriptions...), and training student charges, as well as internal projects for members in a temporary difficulty.

We maintained **LIP6 project** support with the same budget as previously. The goal of this call is to offer inside the laboratory, support for inter-teams projects which, often, lead to proposals at national or even European calls, and also industrial collaborations. Each year around 15 projects are funded. In 2017, we create a new **call for inviting foreign researchers**. They have to come at least two weeks and they must deliver a

seminar open to all the LIP6 members. LIP6 takes in charge transport and accommodation. These two calls are managed by the scientific council. In 2017, LIP6 funds 8 invitations and 15 LIP6 projects.

In 2014, we proposed to the laboratory council to vote a management tax over European or direct industrial fundings. The amount is of 5% and it is used for common action charges. It was first applied in 2015, and , the first income was received in 2016. This budget supports our **incitative policy for applying to European call or direct industrial collaborations**. Indeed, we hired a **project manager** who is in charge of monitoring calls and informing regarding the different opportunities of fundings. She organized several information meetings or industrial partnership meetings... She is involved in the set-up of projects in collaboration with researchers, and when the proposal is sufficiently advanced with the DGRTT (direction générale de la recherche et du transfert de technologies).

We initiated an incitative policy for applying to ERC (European Research Council). With Marc Shapiro, our project manager and the help of ERC successful candidates, we prospect for possible candidates, we examine their projects and we try to encourage them to submit a proposal. We use also all the supports offered by UPMC and CNRS, and eventually we use the expertise of some external SME.

Born in 2012, at the initiative of the actual director, the **colloquium informatique de l'UPMC-SU** works very well, with 5 to 7 talks per year. Invited speakers are very famous researchers some are owner of prices like Turing award or Gödel prize. It is always a successful event, open to all the scientific community and sometime more (i.e. Richard Stallman or Isabelle Collet). The scientific committee of the colloquium is now composed of 12 members, we were 5 at the beginning. It is now completely integrated in LIP6 life.

Administrative and financial service (SCA). This service is currently composed of of 11 financial administratives (2 CNRS and 9 UPMC). It is managed by the secretary general of the laboratory, Ms Aurore Marcos. Each administrative agent manages his/her own portfolio of contracts belonging to one or more teams of a same department. We try, as much as possibly, to share uniformly the works with respect to maintaining a local service, with at most one administrative agent by team. For each contract, they manage the financial part as well as the administrative one.

Since the beginning of our mandate, this service has suffered of many changes. We lost unfortunately and sadly two colleagues due to health issues, which deeply affected all agents. One was replaced very quickly with an internal change of job within UPMC. For the other, a position was offered by the university but the selected candidate did not take the position. Hence, we are expected someone in 2017, before the end of the year. Moreover, we had three departures for retirement, two are already replaced, and for the last one, a recruitment process is ongoing by CNRS. Two agents moved for personal reasons, one was our administrative manager in May 2016 and replaced in January 2017, other one was replaced quickly by CNRS. Then, one agent successfully integrated CNRS, left in December 2016, and will be replaced soon by an UPMC internal change. The service is rebuilt step by step after two difficult years and will find soon a reasonable number of agents for a laboratory of this size. We can only thank our current agents who, despite all these changes, maintained an excellent service for the researchers.

Organism	Service	2012	2013	2014	2015	2016	2017
CNRS	SCI	8	8	8	8	9	8
UPMC	SCI	4	4	4	5	5	5
	SCI total	12	12	12	13	14	13
CNRS	SCA	2	3	3	3	1	2
UPMC	SCA	9	10	8	8	7	7
	SCA total	11	13	11	11	8	9
CNRS	COM	1	1	1	1	1	1
UPMC	DIR	1	1	1	1	0	1
LIP6	Total	25	27	25	26	23	24

Research support and informatics service (SCI). Among the thirteen engineers (fourteen until May 2017, retirement), ten administrate the team servers and the network structure of our laboratory. Two of them, manage the data base of the laboratory, and two others are CSSI (information system security correspondent *correspondants sécurité des systèmes d'informations*). These engineers are dispatched in three groups taking care of three teams' groups: three engineers manage the servers of DAPA-DESIR departments, four are sharing the management of those of NetSys-SysComp-CalSci, and two are in charge of the SOC department. Moreover, these

engineers are involved in the maintenance of Calculus servers (GPU and CPU clusters) and software servers mainly used by the teams they are in charge of.

Pooling server services, like mail, backup and/or account servers with machines belonging to the laboratory is always a topical and hard subject. In June 2017, we obtain a small progress. The directors asked to the managers of DAPA DESIR departments to give their mail server to the laboratory. They accepted. This will allow to host all the mail services of all the researchers on a same server. Hence, we could buy a backup server for the laboratory, as some of the teams became obsolete. Next pooling step will be to purchase an accounts' server. We expected that since the beginning of our mandate in 2014.

Then, three engineers are dedicated to experimental research platforms, like the testbed of FIT or the electronic platform. These engineers administrate these machines and take part of the research projects linked to FIT.

From these thirteen engineers, nine belong to CNRS (two research engineers, seven Engineers), and four have the UPMC status (two research engineers and two engineers).

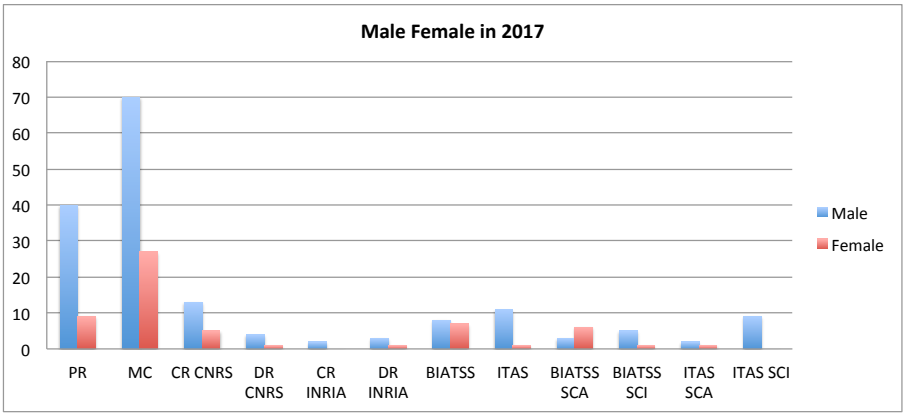
In 2014, we created an **informatics service/ users committee**, which deals with the users demands, and participates to the technical option selection, as well as the annual budget demand for the common services. This committee is composed of about ten researchers and the engineer team.

Parité

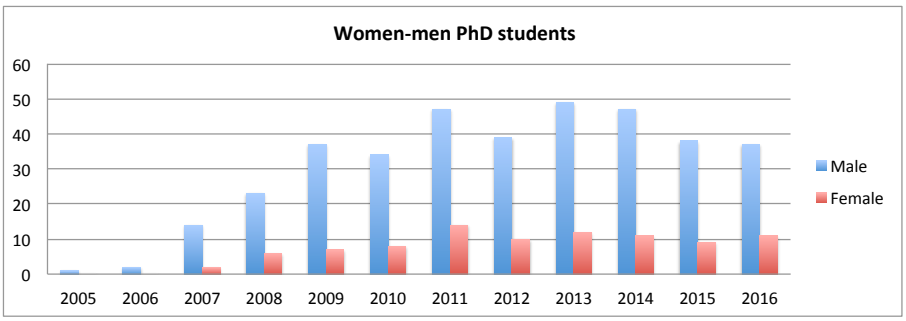
The graphic given below, gives the number of women and men in each categories of the permanent staff of our entity. We can note the the ratio of women for the assistant-professors and the junior CNRS researchers is close to 28%, while it is of $18,5\%$ for the professors, but 20% for senior CNRS researchers. The number of INRIA researchers (6) is too small to be considered. About the agents of the services, the ratio represents $46,5\%$ for UPMC and only $8,5\%$ for CNRS. We can explain this difference by the fact that most of the CNRS agent are technical engineers compared to UPMC agents who are administrative-financial agents (in this case 67% are women).

About the assistant professors and professors, we should expect a progress due to the composition of the recruitment committee where at least 30% are women. Furthermore, this is dependent to the ratio of women in the pool of candidates.

Regarding agents, we note the linkage with the work and the graduation. Most of the women are technical agents, when men are engineers. We do not have any women in the informatics service, which is really not acceptable but not easy to change. We do not have any significant changes in the engineer team.



Regarding PhD students, since 2012 the ratio of women is of around 20% with some variations, but between 19% and $20,5\%$ depending on the year. We note that this ratio is lower than the one of the assistant-professors and the average taking into account by the professors.



Thus, the ratio of women of PhD student really needs to be improved, if we want to change this. But it is not only with the recruitment of PhD student that we must act, we must do it at the master's level degree too.

Protection et sécurité

LIP6 follows the strategy developed by the DGRTT of UPMC about the protection of the potential for use. All cooperative work generates an agreement between the different organisms (public or private) involved in the project. In this document, expected returns are established to fit with each partner?s contribution, in particular with intellectual property.

For integrating the LIP6, people not belonging to our main partners (UPMC, CNRS, INRIA) must set up an hosting agreement between UPMC and their own institution. This applies to a full member of the LIP6, as well as for a training period, or to be an associated member.

Access to the laboratory are under control of an electronic key checking (Kaba). A key is given after the validation of the hosting agreement and notify in our data base.

Opening of an account on our servers, follows the same process, and can be realized only when all the administrative duties are done.

Publication signature policy follows the rules given by UPMC while respecting all structures rules in which a researcher is involved (UPMC, CNRS, INRIA, others). Morever, all the LIP6 publications are now entered in HAL archive and extracted from HAL to the LIP6 data base.

Members of LIP6 are regularly made aware on the security rules, in order to protect themselves against any robbery, as well as computer intrusions. The laboratory has a firewall infrastructure, different levels of wifi access, separated accounts for mail, web, and servers. Two engineers are CSSI (security corespondent): one is UPMC, and other CNRS. They inform users of the different risk and they signify any trouble to the RSSI (central security of UPMC). We take advantage of the UPMC or CNRS services which warn us in case of any IT security failure. Also we receive warnings from the CERT-FR (French security center) and the ANSSI (IT national security agency).

4. Analyse SWOT

STRENGTHS

- LIP6 is well known partly due to its large scientific spectrum which ensures to tackle complex research projects combining both theory and application, in particular with respect to the mains axes defined by the scientific policy. LIP6 takes advantage of UPMC reputation and Sorbonne University actions.
- Numerous collaborations are made in our entity, in particular with industrial partners. However, half of the resources comes from academic organizations, French and European.
- Our researchers are recognized for their scientific excellence with three ERCs, three chairs of Excellence, and seven members of the University French Institute (IUF) during the period 2012-2017.
- Main research activities are integrated in Master cursus which participates to attract about 400 graduate students every year from many countries. Some of our colleagues teach within the MPRI (*Master Parisien Recherche Informatique*) that is common with prestigious schools, namely ENS and Polytechnique. Our laboratory represents more than 50% of the PhD students registered to the doctoral school EDITE.
- Main axes of LIP6 clearly fit the national research strategy (S.N.R.). This program defines the major research topics for the French policy – please refer to https://cache.media.enseignementsup-recherche.

gouv.fr/file/Strategie_Recherche/69/3/rapport_SNR_397693.pdf. More specifically, LIP6 research axes and contributions are in perfect cohesion with Challenge #7 of the France Europe 2020 report – https://cache.media.enseignementsup-recherche.gouv.fr/file/Strategie_Recherche/26/9/strategie_nationale_recherche_397269.pdf.

- Since 2014, we initiated an incitative policy for submitting European projects to Europe and for setting up of industrial partnerships. The policy is supported by the DGRTT and the arrival of an internal project manager. We plan to hire more project managers, to have a team support matching the size and the number of projects.
- Researchers benefit of the different committee actions in order to work in good conditions with an efficient management of the space, a support for publication, LIP6 projects, ...

WEAKNESSES

- As LIP6 is a huge laboratory, it is quite challenging to ensure the visibility of all the diverse research work developed within the laboratory. We bring out three mains axes which cover almost our research. But this identification is recent (2015), we now have to work at the team level, which needs to refine its research domain according to these three main axes.
- LIP6 is now organized in thematic departments, which looks like a partition of the laboratory hiding existing internal collaborations between the teams and the link of them with the main axes.
- Grants from UPMC and CNRS represent less than 10% of the total resources of the laboratory. A big part is used for the common technical support. Hence, the amount for scientific actions is limited, which makes difficult to maintain a coherent scientific policy.
- Financial support of a team is mainly due to its successful projects. The three main drawbacks of this situation, are (i) a big resource disparity between the teams, (ii) sometimes big variations of resources for a team, and (iii) an opportunistic research which does not alway matches the priorities of the laboratory. This makes it challenging to manage a scientific policy for the laboratory.
- Industrial partnership are quite difficult to build. Most of the time, difficulties are due to legal constraints. French Law and organization internal rules drastically slow down the contracting process. Sometimes, this impedes hiring and future collaborations.
- The number of applications to ERC is very low despite all of our efforts. The main reason is the lack of time, in particular for assistant professors and professors. It is also the case with application to IUF. We are currently trying to tackle this issue with a support program.
- The decrease of the number of assistant professors and the increase of the number of students become dramatic for our research. Despite that, results are maintained at their level, but the price to pay is high and we do not know for how long we will maintain such a level with decreasing resources.
- The size and the scientific broad spectrum of LIP6 make decision processes cumbersome. Furthermore, the partitioning into departments makes it complex with some temptations of withdrawal which can prevent the evolution of the laboratory.
- Inside the support services, which are important for an efficient functioning of the laboratory, variation of the number of agents or the behavior of some, can have dramatic consequences first for the members of these services, but also on the research activities. For instance, we stayed nine months without a general secretary (staff and services manager). This weakened us a lot. Moreover, informatics service is still too dedicated to the teams, it does not have a laboratory spirit.

OPPORTUNITIES

- Most of the teams have a strong network for establishing collaborations at different levels: national, European, and international. This network is strengthened by the help of our project manager, the agents of the DGRTT, and the European affairs office. We will continue to encourage, namely for European H2020 projects.

- Similarly, LIP6 maintains a good network of industrial partners and contacts. Currently, LIP6 has 55 running industrial partnerships. Most of them are CIFRE PhD grants. These industrial partners are often involved in collaborative programs and in different calls. This network gives us the opportunity to be very reactive. Most of these actions are supported by "competitive clusters" (*i*.e., *pôles de compétitivité de la région*) or by "*collectivités territoriales"*.
- LIP6 researchers are often required to contribute as experts by numerous bodies like IRT (Research and Technology Institutes), competitive poles, ANR, ANRT, Co-CNRS, CNU... Through these expertises, researchers have a good knowledge of all those structures.
- As we are involved in several LabEx, Sorbonne University Institutes, and regional institutes: DIM (RFSI, PRADA, SIRTEQ), LINCS (about network), PCQC (Quantum Computing); IRILL (free software)... LIP6 is a major actor in the development of IT research in Ile-de-France. Furthermore, we actively participate to the creation of the Tremplin Carnot named "*Interfaces*", which is a good opportunity to meet new potential industrial partners.
- Our courses have good national and international recognition. They help to attract many good students from different origines, cursus, or country. Training periods in industry frequently generate a CIFRE proposal. Master students and doctors are highly headhunted. In Région Ile-de-France, the employment market is lively in Informatics.
- For all the technology transfer actions, we have the help of UPMC (DGRTT) and SATT Lutech. These services contribute to the creation of start-up, patents, softwares,...

THREATS

- During the period, we had to face the departure of around thirty senior assistant professors and professors. At least ten of them were not replaced at UPMC (five effective departures, six in secondment or standby). This has two direct effects: a difficulty to maintain a knowledge, and an increased workload in teaching for the remaining staff.
- The uncertainty on long term resources, also has an influence on the research quality. Researchers are always looking for new funds to prevent any decrease. For some teams we have an excess of projects with sometimes some difficulties to produce the requested deliverables, and for others, a lack of funding.
- The Attractiveness of Saclay development could overshadow our current partnerships, for example: the moving of Telecom ParisTech, the concentration of the CEA on Saclay,... We must stay vigilant, even if this does not have any effect in our work yet.
- In front of the international competition, with less attractive salaries, we have some difficulties to convince well known researchers, or excellent post-dotoral researchers, despite the attractiveness of Paris, to join us.
- Regarding the management of the laboratory, we tend to have a conservative attitude despite the need of evolution of our structure.

5. Projet scientifique à cinq ans

This section presents the project for LIP6 during the period 2019-2023. First, we explain how we analyzed the current situation among the research teams, then we detail our scientific project before sketching its implementation.

1 State of the art and methodology

LIP6 is a large laboratory where a wide range of topics in Computer Science are studied. It is currently partitioned into 6 departments, gathering 22 teams and three research axes were outlined to represent the laboratory. As each of the departments operates differently, handling such a diversity while targeting the emergence of a collective vision for LIP6 is not an easy task. The first step to build the LIP6 project was to elaborate a "state of the art" by visiting all the teams. The goal of these discussions was to get a better vision about everyone's work in order to propose a reasonable basis for discussions that encompass all topics of LIP6 while fostering

the emergence of new collaborations. This process led to setting up a dashboard composed of one data-sheet per team, together with a relation graph between teams from several perspectives. This graph collects common publications (176 from January 2012 to June 2017), projects (funded by LIP6 on its own funds or externally by official agencies or industry), existing relationships and potential topics for collaboration (both exhibited during the discussions).

Figure 1.1 shows one of the most interesting parts of the dashboard: the graph representing potential collaborations. Nodes represent teams (code colors corresponds to current structure into departments) and arcs relationships. Arrows show which team mentioned this potential collaboration (sometimes, both did).

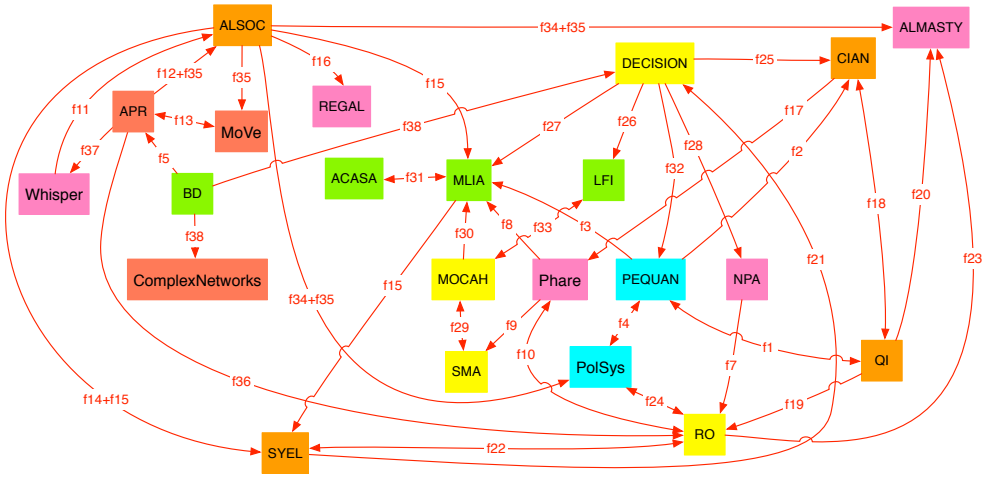


Figure 1.1: Potential scientific cooperation between LIP6 teams. Labels on the arcs refer to Table 1.1. A similar graph (not presented in this document) depicts existing collaborations.

f1 HPC for quantum computing

f2 precision of fixed and floating point calculus + generation of corresponding hardware architectures
f3 computing of numerous small linear systems
f4 certified computing (symbolic numeric calculus)
f5 Management of graphs of requests for experiments
f6 execution of "top-K" requests, use of multi-criteria techniques to enhances results
f7 optimizations in networks
f8 "machine learning" modules for automatic control of networks
f9 intelligent agents for the automatic control of networks
f10 optimization of quality of service and dimensioning in networks
f11 Single Instruction, Multiple Data (SIMD) programming
f12 on parallel computing
f13 on the complementary aspects of verification
f14 hardware implementation of decision trees
f15 hardware implementation of artificial neurons
f16 on "lock-free" synchronization and data-dependent new algorithms
f17 RF networks and mobility
f18 quantum simulation (possibly on FPGA) + security for quantic transfers
f19 quantum algorithms
f20 on cryptography
f21 multi-criteria decisions to set-up non functional characteristics of embedded systems
f22 management and optimization of constraints
f23 problem resolution for crypto-systems attacks (e.g. subset sum)
f24 optimization, Lassère hierarchy and semi-defined programming
f25 optimization for computer aided design (CAD)
f26 optimization for networks
f27 learning of preferences for multi-criteria decision
f28 optimization in networks
f29 virtual agents in serious games + modeling and simulation of systems in teaching + analysis of user's behavior
f30 analysis of textual exchanges between a MOOC participants
f31 complementary approaches for automatic learning
f32 GPU calculus for inference on graphical models + finite precision calculus
f33 advanced clustering techniques for educative data + detection of emotions when learning
f34 cryptographic codes and analysis of physical attacks
f35 security of applications implementation + formal methods for robustness
f36 algorithmic on graphs et combinatorial games
f37 relations between analytic combinatorics and type theory
f38 large graphs and databases

Table 1.1: Topics identified in the graph of potential interactions shown in Figure 1.1.

This preliminary work gives a global vision of potential collaborations inside LIP6. It also shows some relationships that can be enhanced and encouraged in order to propose new original solutions to the challenges LIP6 tackles. The project described below reflects our current vision on our strategy for the coming years, based on the work described above. This vision is still under progress and next year, several meetings and working groups will be created to refine it.

2 Scientific project

Over these past years, the LIP6 has contributed to the emergence of new ideas and methods dedicated to tackle challenges raised by the important needs of :

- digital economy to increase both the security and reliability of computer systems and analysis of data, the development of network infrastructures, the rise of smart cities and other kinds of ambient distributed systems, etc.
- our modern society which expects us to solve privacy and ethical issues, unfair use of social networks, new approaches to human learning, etc.
- and, in other sciences, to use high-performance computing and to exploit big data (e.g. bio-informatics, climatology, cyber-physical systems, quantum systems, etc.) as well as digital humanities.

The rapid growth of these needs and their evolutions will gradually emphasize limitations of existing solutions. In particular, they may fail to scale up and/or suffer from a lack of reliability. This is a source of new challenges for our laboratory, which encompasses a wide area in computer science.

Our objectives are to develop a research activity towards a shared goal of providing end-user solutions that scale up, are safe, secure and reliable, on the range of the future challenges we intend to tackle. This requires the development of new theoretical tools allowing to deal with the intrinsic complexity of our problems that most of the techniques we develop need to overcome, and the new dynamicity of data and digital objects. To do so, we need to exploit the core interactions between hardware evolutions, new computing capabilities and algorithms, as well as the exploitation of huge amount of data.

Hence, LIP6 aims at being a place where cross fertilization between fundamental research and high-level application skills leads to innovative solutions transferable to industry and to other sciences. LIP6 also aims at being a place where research and knowledge anticipate the future needs of science, society and digital economy.

Anticipating these upcoming challenges requires developing models and abstractions as well as efficient theories. Models and Theories need to be versatile enough to be applied in various contexts. All teams at LIP6 share the same approach: *(i)* develop appropriate models and modeling approaches, *(ii)* build theories on these models to produce innovative methods and tools *(iii)* analyze the results obtained on applications to improve *(i)* and/or *(ii)*.

We are convinced that many upcoming challenges for computer science will be of interdisciplinary nature. These challenges will yield scientific issues to computer science as a scientific discipline. Tackling them through these issues will require to develop new methods exploiting the whole span of the LIP6 expertise. The teams of LIP6 develop computer science methods for a wide variety of application domains. While it is important to pay attention to all these application domains, we have identified three on which almost all the teams of the LIP6 may contribute and that are related to initiatives supported by our University.

- *E-health*. Computer science is a major tool for improving both quality and costs of medical acts and medicine systems, as well as preventing diseases and predicting the behavior of patients. For this, we need to develop systems and numerical tools for to help decision making, diagnosis, analysis of a huge amount of data collected by sensors thanks to IoT based-solutions and to develop cyber-physical systems, while protecting personal data.
- *Smart cities*. Computer science can offer solutions to optimize costs, improve the organization of cities while ensuring the well being of inhabitants. There is a need for using smart devices to measure various phenomena such as air pollution and noise pollution. Such devices collect a huge amount of data coming from many sources, that need to be processed and protected since their diffusion may alter privacy. IoT-based solutions also play a prominent role and open many perspectives for the development of smart cities while, again, privacy issues in this context are critical.
- *Ethical issues and humanities*. When the number of computerized systems increases, ethical and human issues rapidly arise to define appropriate usage of such systems. As an example, what should be decided in an automated vehicle when lives may be affected: protect the passengers, give priority to other users?

or try to reduce the number of casualties? such problems are easy to extend (but not to solve) when humans and automated systems interact. In these areas, there is a need to dig for answers from the social sciences. Similar issues exist for assisted human learning where connections obviously exist with didacticians.

Beyond these applications, a role is to increase the impact of computer science in other sciences. This is consistent with the scientific policy of UPMC and its pole "Modélisation et ingéniérie". As such, well-identified applications of first importance will be investigated, in particular towards sciences such as biology, climatology, physics and mechanics.

Scientific themes. To deal with these Computer-Science challenges, we need to continue our research activities in the heart of the computer science discipline. The following four key scientific themes which cover the research in LIP6, and which will be used to guide management, are as follows :

- *New era of artificial intelligence: data science and intelligent systems.* As already mentioned, data analysis, adaptive learning and reasoning, as well as decision making and intelligent cooperation between agents play a key role in the challenges LIP6 tackles. Our goal is to push forward the current state-of-the-art to : *(i)* elaborate techniques and tools that scale up enough to increase the impact of these techniques in other sciences and, *(ii)* enter in the era where decisions/recommendations made by AI to end-users (such as executives, customers or teachers/learners) are justified.

- *Next generation of networks, architectures and systems: handling dynamicity and autonomy.* Future systems will rely on strong execution infrastructures (network + operating systems) able to support new features like dynamicity, mobility, or real-time. Moreover, energy consumption and quality of service are issues to be considered too. Finally, costs (conception, deployment, maintenance) have to be considered in situations where availability is crucial and where the execution of such systems is distributed over devices of different natures (from smart cards to large computers).

- *New approaches for safety, security and reliability of software and/or hardware systems.* More and more delicate missions are being handled by automated systems : life-critical (e.g. autonomous driving or aeronautics), mission critical (e.g. space missions), business-critical (e.g. e-government or business systems). It is thus crucial to ensure that such systems behave as expected, which is difficult when, in parallel, their complexity increases, due for example to highly parallel execution, the existence of failures or attacks, etc.

- *New frontiers of mathematical computing.* Mathematical algorithms and their high-performance implementations are crucial for solving the problems appearing in the applications we target. We need such algorithms to optimize computer systems, devices and decision making, to increase the reliability of computing software and/or to assess the security of cryptographic primitives and prepare the quantum era. Most of them suffer from the curse of dimensionality and our goal is to push forward the current frontiers of mathematical computing.

At this stage, these themes still need to be refined and their related goals may evolve thanks to future discussions. Indeed the best structure to organize research activities at LIP6 is still under construction.

Interdisciplinarity. On many fundamental aspects, research at LIP6 is pluri-disciplinary and typically involves the use (and definition) of mathematical tools to model and solve our problems. LIP6 also intends to develop strong connections with other disciplines where it can increase the impact of computer science to solve their problems.

- **Mathematics.** Many of the models we develop are based on mathematical tools such as statistics and probability (learning, decision theory, algorithms on graphs), combinatorics and combinatorial optimization (average case analysis of algorithms, operations research), algebra (cryptography and symbolic computation) and numerical analysis (optimization, computer arithmetic).

Hence, several teams at LIP6 have close connections with colleagues in the department of mathematics at LPMA[1] (probability, statistics and high-performance computing), LJLL[2] (numerical algorithms and high-performance computing), and IMJ[3] (combinatorial optimization, algebra and number theory).

- **Life sciences and climatology.** These areas are faced with the challenge of analyzing huge amount of data and the need for high-performance computing. Hence, the core computer science methods which are relevant in this context include artificial intelligence and high-performance computing. So far, LIP6 mainly collaborates with the LOCEAN[4] laboratory on these topics.

- **Medicine.** Computer science can bring new methods and techniques to be used in medical devices and/or improve diagnosis and medical acts. These encompass artificial intelligence techniques integrated in devices (disease detection) or help diagnosis or improve medical imaging techniques. Moreover, within the context of the Carnot interface, we could participate in actions to assist in very fine surgery (partially performed by robots) or distance control of surgery (where actors are distributed in several places).

- **Electronics.** LIP6 also has strong contacts with electronics. Some of its members come from that area (CNRS section 8) and our laboratory is connected, as a second institute, to INSIS[5]. This results in an active participation to the GDR SOC2 (ex SoC/SiP). Some colleagues are also active in the IEEE CAS[6] section.

- **Physics.** Most of the contacts we intend to develop towards physics are related to quantum. The goal is to strengthen the development of the quantum era in computer science in close cooperation with quantum physics. The main contact inside UPMC on this topic is the LKB[7].

- **Humanities.** Interactions between humanities, social sciences, and computer science are increasing. They include methods from artificial intelligence for text, image and interaction analysis, as well as graph theory for community detection. LIP6 is involved in several research initiatives fostering research at the interplay of these topics.

3 Implementation of the project

3.1 Governance

By January 2019, the director of the lab should be Fabrice Kordon. He would be helped by two adjunct directors ; Mohab Safey El Din is one of them, the second one is not identified yet.

The direction team. It is composed of the director, the two adjunct directors, and the general secretary.

Strategic decisions are made by the director. However, it seems crucial that the adjunct directors have distinct scientific profiles different from the one of the director, to confront different scientific views on issues to be solved during the term. Adjunct directors help the director in some of his administrative duties related to the lab management, such as yearly meetings with technical staff. They could also handle some tasks such as the relationship with the Carnot.

LIP6 is represented by its director. However, when the director is not available (he still has 64 hours of teaching per year), one of the adjunct director may replace him. A project management environment has been set up to ensure information sharing and to allow collaborative work.

Once a year, the direction team will meet with all teams in order to update our vision of research at LIP6 (via an update of the dashboard) and discuss highlights and any matters that the researchers could raise.

The direction board. It is composed of the direction team and representatives of the four themes. It should meet twice a month to prepare strategic issues to be proposed to the scientific board. The direction team will

[1] Laboratoire des Probabilités et Modèles Aléatoires
[2] Laboratoire Jacques Louis Lions
[3] Institut de Mathématiques de Jussieu Paris Rive Gauche
[4] Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques
[5] Institut des sciences de l'ingénierie et des systèmes
[6] Circuits and Systems
[7] Laboratoire Kastler Brossel

also consult this board is about current affairs of LIP6 (urgent demands, management matters, etc.). When needed, external experts, or other members of LIP6 will be invited to discuss a specific topic.

The scientific board. The scientific board has an advisory role. It is composed of the direction board, and one representative per team (the team leader or any other member mandated to represent the team). It meets about once a month upon an invitation from the director and handles all scientific concerns of the laboratory. If needed, external experts, or other members of LIP6 can be invited to discuss a specific topic.

The laboratory board. This board is organized after the statutes we inherit from the CNRS as an UMR. The direction team is ex-officio member of this board which also gathers 20 members. 12 are elected (4 representatives from the "collège A" – research directors and full professors –, 4 representatives from the "collège B" – researchers or associate professors –, 2 representatives from the "collège C" – administrative staff and engineers –, and 2 representatives from the "collège D" – Ph.D. students and pots-docs). The other members are nominated by the director.

This board meets about once a month upon an invitation from the director. Typically, meetings of the laboratory board will take place about one week after the scientific board, so that information on scientific issues needing to be also discussed there could benefit from the input of all the members of LIP6 (via the laboratory board members).When some topics are of interest to both the scientific board and the laboratory board, these two boards may meet together.

Research teams. They are the ground-basis for research. They are represented by a team leader who is nominated at the beginning of the term by the director after advice from team members and the scientific board. There are currently 22 teams in LIP6. We would like to establish a procedure to create new teams (as for Inria) and to terminate a team so that, changes of configurations become easier at this level. This should enforce the flexibility of configurations at LIP6.

Scientific themes. The four themes have two objectives. They represent a forum where scientific issues should be debated and skills crossed in order to imagine new solutions to push forward the limits of our discipline. The idea is to take advantage of having various competencies inside the lab and generate discussions among its members.

Of course, governing procedures have to be defined for these themes, which will be the subject of further discussions before the new direction team starts its leadership. It is important to us that a minimum set of requirements are explicit so that a common understanding (and way of management) is possible.

3.2 Scientific animation

Scientific animation starts in research teams, under the impulse of team leaders. The work inserted in transversal actions is followed up at the level of one or more themes. The objective is to enhance interactions and cooperations between LIP6 members and research teams.

The direction team will allocate resources to emphasize this communication via mechanisms inspired from the current "LIP6 projects". Such fundings will be allocated by the direction team after advice from the scientific board, but proposals could be emitted by teams or by themes. We also aim at organizing yearly "LIP6 days" gathering all members and presenting some focus on the work done at LIP6, possibly around the scientific themes.

Scientific animation is also important at the level of our Ph.D. students. So, Ph.D. seminar should be encouraged at the Themes level, and some of the most interesting presentations, pushed forward to a presentation at "LIP6 days". We would also like to set-up a training of our students to prepare their application for academic positions, with possibly some public training to present their work as in an audition.

Similarly, it is important to support arriving colleagues with some starting budget, allocated during their two first years at LIP6, allowing them to buy their computer and to present their results in conferences before they are involved in research grants and projects. The existing working group encouraging and helping colleagues in their application to research grants like IUF or ERC will be reinforced. This could be done in relation with similar mechanisms and grants when established by our university and the CNRS.

3.3 Human resources

Faculty members (researchers, associate professors and professors). As sketched by the above project, in

order to develop its scientific strategy, LIP6 needs to cover a wide span of expertise. Currently, because of its size, LIP6 looks like a giant and the total number of faculty members seems to be stable. Actually, it turns out that the situation is more complex. While some groups have been created from scratch or were welcoming colleagues, other groups have lost faculty members. All themes have been impacted, in particular, several groups working on artificial intelligence, systems and mathematical algorithms.

In this context, we need to encourage mutations to LIP6 of CNRS researchers to compensate for the loss of faculty positions. As faculty positions are becoming rarer, the top priority is to ensure the highest scientific quality of the hired people. We plan to rethink our recruitment policy by defining position profiles on a pluri-annual basis. This is one way to be reactive on the quality of the candidates while a large number of topics is present in the laboratory. This also requires us to identify pools of potential candidates rather early in the process so that profile proposals sent to our institutions are sound and will thus gather good candidatures.

PhD Students. The attractivity of research to master's students is hindered by the high salaries and the strong demand for computer scientists in industry. In a context where our number of HDR grows, we need to attract more Ph.D. students. Thus, we would like to support any initiative in that could encourage good students to come, for example from foreign countries. This could be achieved thanks to research collaborations leading to co-supervision and "co-tutelles".

Administrative staff. Given the amount of contracts the LIP6 is involved in, it is important to have efficient administrative support. In its current organization, preferred by our staff, administrative people handle all the work needed to manage teams (missions, contracts, purchases, etc.). This allows them to have a broader activity but load issues when the amount of work change (e.g. due to a variation of contracts) should be handled to maintain a reasonable work balance among all staff members.

Engineers. It is important to increase the sharing of services in our laboratory. This will save human resources that could be redeployed towards support activities to research. Such activities would be driven by the laboratory, after advice from the scientific board ; in particular, the profile of our engineers is appropriate for the definition of experimental platforms (e.g. GPU platform) that could also be shared with other entities when it is appropriate.

3.4 LIP6 at Sorbonne Université

LIP6 has the potential to play a structuring role in our discipline in our environment. For instance, we are involved in several actions in the UFR d'ingénierie, our university, and in Sorbonne Universités.

In the UFR d'ingénierie, the Labex SMART and the tremplin Carnot Interface federate several laboratories and offer a strong opportunity to develop and to valorize our activities, especially in a pluri-disciplinary context (*e.g.* e-health, augmented reality, ambiant intelligence). Similarly, activities could be conducted with the new "Institut Universitaire d'Ingénierie en Santé", in which the labex Smart is involved.

In the future Sorbonne Université, LIP6 is involved in the Labex CalSimLab with several applied mathematics laboratories. We also collaborate with several laboratories dealing with data sciences or high-performance computing. These are applied to climatology (with LOCEAN), physical sciences (LKB) or biology (Institut du Cerveau et de la Moelle Épinière). The LIP6 is also involved in the labex Obvil which establishes a link between computer science and humanities ; such interactions are original and should be encouraged.

Beyond Sorbonne Université, LIP6 participates very actively to several clusters such as Cap Digital, Systematic, Advancity, Medicen, etc. thanks to its involvement in applications. It will for sure keep on maintaining and developing this strategic activity.

DEPARTEMENT **CALSCI**

1. Présentation du département

The Scientific Computing Department encompasses research activities related to scientific computing at LIP6. The mainstream objectives consist in conciliating high-efficiency with numerical reliability of computing software. Reliability issues are tackled through the development of mathematical algorithms relying on

— *computer arithmetics* for rigorous floating point computations in numerical linear algebra, approximation of special functions, etc.

— and *symbolic computation / computer algebra* techniques yielding exact results for e.g. mathematical problems over finite fields or non-linear problems over continuous domains.

In both contexts, efficiency issues are tackled through the development of high-performance computing paradigms and libraries freely distributed to foster a high impact of our contributions and the transfer of our formal techniques to industry.

The mathematical algorithms developed there bring scientific foundations to many areas of computer science, in particular,

— cryptology : the security of crypto-ciphers can be modeled through mathematical algorithmic problems usually expressed in finite fields ; in this context algebraic algorithms provide the foundations of security analysis.

— reliable high-performance computing : making high-performance reliable from the numerical point of view is a deep challenge ; in particular ensuring numerical reproducibility of high-performance computations becomes a strategic goal for many high-tech industries.

Other topics such as signal processing, computational geometry and imaging are also addressed.

The structure of the department in two teams, PEQUAN focusing on computer arithmetics and validated numerics, and POLSYS, joint team with Inria Paris Center, focusing on polynomial systems through algebraic computation, reflects these two computational methodologies.

The overall number of faculty members in the department has decreased in both groups during the last five years despite the increase of the span of the successfully investigated applications, the increase of their respective activities, and the initial small size of the department.

2. Politique scientifique

The key feature of the department is to ensure reliability for mathematical software. Our analysis and shared belief is that transferring the methods we develop either to other academic areas or the industry requires the development of highly sophisticated software. Our strategy to increase the impact of our results is then to develop and make available computing libraries, freely distributed, and providing the top state-of-the-art algorithms for the problems we intend to solve. With that in mind, we emphasize that in order to ensure sustainability of the produced software, these are mostly produced and maintained by permanent faculty members.

Some of these software are now interfaced with others developed e.g. at Laboratoire Jacques Louis Lions and the development of others has been supported through projects funded by the LabEx CalSimLab and Institut des Sciences du Calcul et des Données. This has created the opportunity for the department to foster collaborations with the Laboratory Jacques-Louis Lions (LJLL) and Laboratoire de Probabilités et Modèles Aléatoires (LPMA) in the Mathematics department of UPMC. Beyond our intrinsic common scientific interests, these collaborations make easier the transfer of our technologies to the mathematical communities which have

strong connections with industry. They allow us also to become the preferred contact on all activities requiring more reliable computing software on the range of problems we currently study.

To encourage these activities, the department animates the joint seminar on interactions of Computer Science and Mathematics at UPMC. Several projects and connections have been raised through this seminar. The department is also involved in the administration of joint curricula between Mathematics and Computer Science both at bachelor and master degree.

Finally, we emphasize that the contributions of the teams of the department are evaluated through their ability to solve applications in which we have naturally developed a local expertise, e.g. in cryptography, computational geometry and imagery. This has led us to encourage transfer activities either through patents and SATT Lutech (e.g. thanks our research activity on cryptography) or with partnerships with large companies (e.g. Safran and Dassault Systems).

EQUIPE PEQUAN

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The PEQUAN team (Performance and Quality of Numerical Algorithms) is a member of the Scientific Computing department of the LIP6. It is composed of 11 permanents (3 Professors, 6 Associate Professors (2 with HDR), 2 Junior CNRS researchers (1 with HDR)) and 1 Emeritus Professor. We constantly have around 10 PhD students and one or two postdocs/engineers. Professors mainly come from UPMC but a professor comes from University of Paris 2. Researchers come from CNRS.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	12	11	10	10	10	9
Researchers	0	0	1	2	2	2
PhD students	8	10	10	9	7	8
Others	2	5	3	2	0	2
Total	22	26	24	23	19	21

Grants	2012	2013	2014	2015	2016	2017
International	0 k€	0 k€	0 k€	0 k€	0 k€	0 k€
National	160 k€	185 k€	182 k€	153 k€	120 k€	45 k€
Industry	117 k€	116 k€	93 k€	38 k€	42 k€	34 k€
Others	0 k€	0 k€	0 k€	20 k€	82 k€	82 k€
Total	277 k€	301 k€	275 k€	211 k€	244 k€	175 k€

Politique scientifique

Scientific objectives: The main research goal of the PEQUAN team is to perform numerical simulations **fast** and **accurately**. The main anchor point of the team is the use of fine properties and the development of knowledge in **computer arithmetic** and **high-performance computing** (HPC). Those skills are used to solve problems in a wide variety of application fields. Our focuses are mainly:

- **reliable computing**: design of fast and accurate algorithms in fixed-point or floating-point arithmetic as well as their numerical validation with some applications in signal processing ;
- **high-performance computing**: design of parallel algorithms and performance optimization of scientific codes on different architectures (multicore, GPU, Xeon Phi, grid, etc.) ;
- problems linked with **image assimilation** and **scientific visualization**.

The research group tries to promote its academic results into real-life applications via some collaborations with other academic teams or with industry. Our strategy is to publish in the major journals and conferences of our domain. Moreover, we also develop some software that we made available as much as possible.

Scientific animation: Every two weeks, there is a team seminar on Thursday. The seminars are classic one-hour seminars: a French or foreign colleague is invited and presents a recent research result. If needed, this seminar is preceded by a team meeting. Team meetings are not regularly scheduled but organized when needed

or asked by some people. Moreover lunch, that we try to have together everyday, is also an opportunity to get informed about our lastest research activities.

2. Produits de la recherche et activités de recherche

Bilan scientifique

We present our global achievement over the 5 past years, stating it via the 3 main fields of our scientific policy.

— ***Reliable computing***

Numerical validation: The improvement of computational resources for numerical simulation makes crucial information on the numerical quality of results. The CADNA library (`http://cadna.lip6.fr`) developed in the PEQUAN team implements Discrete Stochastic Arithmetic to estimate round-off error propagation in simulation codes in C/C++ or Fortran. It is based on three executions with a random rounding mode that consists in rounding any result upwards or downwards with the same probability.
The C version of CADNA benefits from several recent improvements: for instance, no more explicit change of the rounding mode is required. This new version exhibits better performance and enables one to control vectorized codes. Fortran binding for this version of CADNA is currently developed. Support for parallel codes has also been improved: CADNA can be used with codes using MPI or OpenMP, and with hybrid CPU-GPU codes using CUDA.
The PROMISE tool (PRecision OptiMISE, `http://promise.lip6.fr`) based on CADNA has been proposed to auto-tune floating-point precision. From a required accuracy on the computed results, PROMISE aims at maximizing the number of single precision variables in a program.
We have shown that compensated algorithms enable one to increase the results accuracy even with directed rounding. Whatever the faithful rounding mode chosen, compensated summation, dot product, and Horner scheme algorithms provide a result almost as accurate as if it was computed with twice the working precision. Therefore compensated algorithms and CADNA can be used in a program both to improve and control the accuracy of results.
We have shown how to perform a dynamical control of Newton's method for the computation of multiple roots of polynomials. Using Discrete Stochastic Arithmetic, root approximations are computed until the difference between two successive approximations is numerical noise. With such a stopping criterion, the optimal number of iterations in Newton's method are performed. Moreover it is possible to estimate in the result obtained which digits are in common with the exact root.
Works on Discrete Stochastic Arithmetic has recently enabled us to develop collaborations with CEA, EDF, IDRIS, INSERM, IRT System X, Laboratoire de Chimie Théorique (UPMC), ONERA, Univ. of Stuttgart, Queen's Univ. of Belfast.

Code generation and reliable *high-level* operations: Computer arithmetic, in particular floating-point arithmetic, is currently reigned by standards, such as IEEE-754, defining basic operations all other code is built on. For instance, the IEEE-754 Standard specifies the exact floating-point behavior of basic operations, such as addition or multiplication. Code for numerical analysis is then built upon these operations. However, the gap between the high-level operation used in numerical analysis and the very basic operation defined in the standard is huge. We work on reducing that gap.
Our work started with code generation for mathematical functions, such as e^x or $\sin$, with guaranteed accuracy. The resulting code is the higher-level operation implementing the given mathematical function. This work is continuing and about to be extended to other domains, such as code generation for linear time-invariant filters.
Additionally, we have worked at extending floating-point arithmetic to high-level operations such as faithfully rounded norms of vectors, correctly rounded operations extending IEEE-754 such as Fused-Add-Add ($a+b+c$) or Fused-Multiply-Twice-And-Add ($a\times b+c\times d$). We have done research with respect to floating-point arithmetic in mixed bases, combining numbers represented in base 2 with numbers represented in base 10. We proposed algorithms for mixed-basis comparisons, computed worst-cases

for mixed-basis operations such as Fused-Multiply-And-Add and designed code for correctly rounded conversion of arbitrary length decimal strings to binary floating-point numbers.
Our work on numerical software frameworks, such as Sollya (`http://sollya.gforge.inria.fr`), allows us to reliably perform high-level operations on even more complex objects, such as computing and validating norms on mathematical functions, computing rigorous polynomial approximations, computing norms on Linear-Time-Invariant filters or validating filters against frequency specifications. For example, we proposed rigorous approaches to compute the Worst-Case-Peak-Gain of a filter and to rigorously validate a filter's frequency response with respect to specifications. This allows our fixed-point code generator to produce reliable implementation of linear filters.

Symbolic-numeric computation: At the interface of computer algebra and floating-point arithmetic, we are working on a long-term project to develop automatic code generators for the implementation of special functions (with first results published in 2015), and occasionally on other topics in floating-point arithmetic. Our work on special functions is based in part on a package for symbolic-numeric computations with D-finite functions in the SageMath computer algebra system. New theoretical results related to the development of this package include yet unpublished algorithms for computing tight bounds on the tails of D-finite series. Its development also was the impetus for substantial contributions to SageMath itself (350+ patches since Oct. 2013). In collaboration with the SpecFun team at Inria Saclay, we recently proposed the first algorithms with controlled complexity for finding closed-form solutions of linear Mahler equations.

— ***High-performance computing***

In the context of high performance scientific computing, our research work has first focused on algorithm-architecture matching. Due to different features in HPC architectures (multicore CPUs, GPUs, integrated GPUs, Intel Xeon Phi, Cell processor), we try to determine, for a given scientific application for which multiple algorithms are available, the algorithm-architecture couple that leads to the best performance results. We have hence managed to improve GPU performance for different applications which did not perfectly match the GPU architecture (in computer arithmetic, in atomic physics and for segmentation in image processing). Conversely, we have introduced task-based parallelism on multicore CPUs in a sequentially efficient algorithm that did not exposed massive parallelism (for a recursive fast multipole method). We have also investigated new hybrid CPU-GPU algorithms (in computer arithmetic), as well as the interest of new HPC architectures like integrated GPUs from a performance and energy efficiency point of view (joint work with Total and AMD for seismic imaging).
Another research direction has dealt with SIMD (Single Instruction - Multiple Data) computing, whose share in the overall compute power of HPC architectures is increasingly important (GPUs, AVX-512). We have managed to modify algorithms and implementations in order to reduce divergence for irregular applications, leading to significant performance gains. This has been done for stochastic arithmetic and computer arithmetic (Table Maker's Dilemma). Improving the regularity of computations thanks to matrix formulation and BLAS routines has also been achieved for a fast multipole method on the Cell processor.

On modern multi-core, many-core, and heterogeneous architectures, floating-point computations, especially reductions, may become non-deterministic and, therefore, non-reproducible mainly because of the non-associativity of floating-point operations. We recently introduced an approach to efficiently compute the correctly rounded sums of large floating-point vectors, achieving deterministic results by construction. Our multi-level algorithm consists of two main stages: first, a filtering stage that relies on fast vectorized floating-point expansion ; second, an accumulation stage based on superaccumulators in a high-radix carry-save representation. We presented implementations on recent Intel desktop and server processors, Intel Xeon Phi co-processors, and both AMD and NVIDIA GPUs. We showed that numerical reproducibility and bit-perfect accuracy can be achieved at no additional cost for large sums that have dynamic ranges of up to 90 orders of magnitude by leveraging arithmetic units that are left underused by standard reduction algorithms.

— ***Image assimilation and scientific visualization***

Recently, we have mainly addressed motion estimation using image assimilation techniques. Motion estimation is designed as an inverse problem, the model describing the motion dynamics. We proposed several applications in fluid flow estimation from satellite images, traffic video monitoring, and more recently, nowcasting of rainy events from radar images. In fluid motion we proposed an original approach to compute divergence-free motion using a sine basis representation of vorticity. We investigated the issue of both estimating motion and acceleration. Determination of acceleration allows us to retrieve forces that are not represented by the model. This can be used to detect abnormally events in traffic video, or to recover geophysical forces (Coriolis and gravity) involved in the sea surface circulation. We also investigated the issue of deformable objects tracking and propose an original approach that both determines motion and segmentation of deformable structures.

Regarding scientific visualization, recent results include practical algorithms for the topological analysis of bivariate data (with the first practical algorithm for the computation of bivariate Reeb spaces of volumetric data) and of uncertain data. A specific effort has been carried out at the boundary between visualization and high performance computing, with the introduction of multi-threaded algorithms for the computation of the contour tree.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	11	9	5	8	9	4	46
Ouvrages	1	2	1	1	0	0	5
Conferences	25	19	14	16	19	7	103
Logiciels	2	2	3	3	3	3	16
Brevets	0	0	0	0	0	0	0

Sélection des produits et des activités de recherche

Voir Annexe 4.

Faits marquants

— Optimization of stochastic arithmetic for HPC applications - [PEQUAN-J-7]

In the context of high performance computing, numerical validation becomes increasingly important because of the higher level of parallelism and of the large number of operations. Stochastic arithmetic, implemented through the CADNA library, estimates rounding errors in simulation codes. However it had a high overhead on execution time, especially for very optimised applications, and did not enable the use of vector instructions. Several optimizations of stochastic arithmetic have been proposed: implicit rounding mode change, new random number generator, new estimation of the correct digits, implicit rounding mode change, inlining of overloaded operators. Thanks to these optimizations, the overhead of CADNA has been reduced by up to 85% on various benchmarks. Furthermore this new CADNA version enables to take benefit of vector instructions. CADNA has also been improved to enable accuracy estimation in shared or distributed memory applications using OpenMP or MPI for communications. This new CADNA version can be downloaded on: `http://cadna.lip6.fr`.

— Revisiting the signal processing implementation problems with the tools and methodology of the computer arithmetic - [PEQUAN-C-2]

In the control and signal processing communities, finite precision errors are modeled as noise and statistically analyzed, but this only gives statistical information about the implementation error (no bound

or interval). Moreover, none of the possible tools to analyze the propagation of the error (modeled as interval) through filters were rigorously computed. During Anastasia Volkova's thesis, classical computer arithmetic tools and methods were used and adapted to signal processing implementation problems.

— Accelerating intensive-computation problems with new algorithms suitable for GPU - [PEQUAN-J-3]

The IEEE 754-2008 standard recommends the correct rounding of some elementary functions. This requires solving the Table Maker's Dilemma (TMD), which implies a huge amount of CPU computation time. We have considered accelerating such computations on GPUs, which are massively parallel architectures with a partial single instruction, multiple data execution. We designed a new parallel search algorithm that is much more efficient on GPUs thanks to its more regular control flow. We also presented an efficient hybrid CPU-GPU deployment of the generation of the polynomial approximations required by the algorithm. In the end, we managed to obtain overall speedups up to $53\times$ on one GPU over a sequential CPU execution and up to $7\times$ over a hex-core CPU, which enable a much faster solution of the TMD for the double-precision format.

— Being both very fast and accurate is possible - [PEQUAN-J-8]

We recently presented an efficient algorithm to compute the faithful rounding of the l_2-norm of a floating-point vector. This means that the result is accurate to within 1 bit of the underlying floating-point type. This algorithm does not generate overflows or underflows spuriously, but does so when the final result calls for such a numerical exception to be raised. Moreover, the algorithm is well suited for parallel implementation and vectorization. The implementation runs up to 3 times faster than the Netlib version on current processors. As a consequence, this shows that it is possible to be more accurate and faster than the state-of-the-art libraries by using the specificity of hardware.

3. Analyse SWOT

— *Strengths:*
The main force of the team is the presence of a dual competence in both computer arithmetic and HPC. There are very few teams that have both if we except James Demmel's team at UC Berkeley and Jack Dongarra's team at UT Knoxville. The team has strong industrial cooperations (EDF, Total, Safran, Intel, AMD, Kitware, etc.) as well as numerous collaborations inside the university (Institut des sciences du calcul et des données, Institut Jean Le Rond d'Alembert, Laboratoire d'océanographie et du climat) and also outside both at national (CEA, LIP (Lyon), IRISA (Rennes), LIRMM (Montpellier), LRI (Orsay), Cristal (Lille), LIRIS (Lyon), INRIA-Grenoble) and international levels (Hamburg Technical University, Waseda University, Queen's University of Belfast, National University of Defense Technology (China), University of Zaragoza, University of Arizona, University of Leeds, University of Utah). Our applications are multidisciplinary ranging from medical imaging to finance. We are particularly active in the development of freely distributed software and libraries (CADNA, Sollya, FiPoGen, etc.).

— *Weaknesses:*
One weakness lies in the fact that members of the team are highly involved in administrative duties both in research and teaching. Two of the three professors have heavy administrative duties (director of IUT Cherbourg, vice-president of UPMC). One other member is the co-head of the track "security, reliability and performance of numerical algorithms" of the master of computer science, one is the co-head of the track "computer vision" of the master of computer science and another one is the co-head of the bachelor of applied mathematics and computer science. One person is also the co-head of the new MAIN (mathematics and computer science) engineer degree at Polytech Paris-UPMC. This involvement in teaching administration of many members of the team takes a huge time and leaves less time to research activities. Another weakness is the lack of engineers and postdocs to deal with the coding part and with maintaining the libraries developed by the team. The workforce on HPC should clearly be strengthened both for teaching and research.

— *Opportunities:*

Some of the most important challenges for the next few years will be Exascale computing and deep learning. Exascale will need more and more precision whereas the challenge for deep learning is to use as less precision as possible. In these domains, the skill of the team could be used in to order to tackle these challenges. Exascale computing will also lead to more and more data to analyse. In-situ analysis combines HPC and visualisation techniques. The team hopes to be a leader in that field. For that, we will try to seize the opportunities for new industrial collaborations and more important to submit EU H2020 projects in that areas.

— *Threats:*
Our research domains need a strong reactivity linked to changes in hardware and/or evolution of standards but also to strong world competition. The need to reinforce our workforce in HPC is crucial: some people specialized in HPC could leave the team soon creating a threat for our ability to be competitive in that domain.

4. Projet scientifique à cinq ans

In the next few years, we plan to work on the main challenges that deal with the problem of accuracy in high-end simulations and with the use of HPC. Concerning accuracy, two different challenges will have to be tackled. The first one is the use of less and less precision in the computation to limit power consumption (green computing) but still having some information about the numerical quality of the result. Moreover, applications like deep learning in big data try to use less precision in order to accelerate the computation and to use less memory. The second one is the use of highly accurate (multiprecision) computation in simulation. The use of HPC (soon Exascale computers) makes it possible to tackle more and more complex and precise problems that need to use huge precision in the computation (for example when simulating chaotic phenomenon).

Moreover, the increasing computational power will lead to more and more data to be produced. The analysis of huge quantities of data will be challenging. Indeed, it will be difficult to store all the data and/or to access them quickly. Visualization of data will need a fast insitu analysis in order to respect time or memory capacities.

A long term challenge will be to able to automatize all those works. As a consequence, we plan to focus our research on the following directions in the next few years.

Numerical validation: One of our perspectives is the design of efficient and reliable BLAS (Basic Linear Algebra Software). We will study how optimizations carried out in the implementation of Discrete Stochastic Arithmetic, such as implicit rounding mode change, could lead to efficient and validated BLAS.

We will work in particular on the design of efficient and reliable BLAS for small matrices. Batched BLAS algorithms have recently been proposed. They consists in multiple independent BLAS operations on small matrices grouped together as a single routine. Recent works focus on the optimization of batched BLAS performance. We plan to study both numerical validation and performance optimization of batched BLAS on several architectures including hardware accelerators.

Numerical validation algorithms should take into account the evolution of architectures and compilers. Solutions proposed to enable efficient numerical validation on various architectures (CPU, GPU, shared or distributed memory supercomputers) should be adapted and combined for large scale heterogeneous environments.

The performance of the numerical validation process could be improved by taking into account information on parts of the code to be controlled (use of methods with a priori error bounds on results, call to correctly rounded procedures).

The link between numerical validation and other challenges of HPC should be studied. Discrete Stochastic Arithmetic that requires several executions and resilience strategies based on replication could be combined. The link between numerical robustness and energy consumption could be investigated, and in particular the impact of precision optimization on energy consumption.

Symbolic-numeric computations: Our main goal is the development and analysis of algorithms and software for "doing mathematics" rigorously on a computer, typically via either exact symbolic computations or numerical computations and error bounds. Two central long-term goals are the motivation for much of our research activity: on the one hand, the development of symbolic-numeric tools for computations with linear

ODEs, on the other hand, that of code generators for the numerical evaluation of special functions. More specifically, questions we are actively working on in the shorter term include the numerical evaluation of D-finite functions given by generalized initial conditions at irregular singular points, the computation of asymptotic expansions of the zeros of D-finite functions (in collaboration with B. Salvy), and an exciting collaboration with S. Chevillard and S. Filip on the analysis of Remez' algorithm for computing uniform polynomial approximations of real functions.

High-performance computing: In the coming years, algorithm-architecture matching and SIMD divergence handling are research directions we plan to further investigate. This will be done for recursive fast multipole methods in astrophysics whose efficient deployment on integrated GPUs is under study. Such fast multipole methods will also be investigated in the context of molecular dynamics (ExaPol project, CNRS InFIniti 2017). Task-based parallelism is also currently investigated for a key algorithm in scientific visualization (contour tree computation). Finally, P. Fortin will be for a one-year CNRS sabbatical leave at Université de Lille 1 for 2017-2018 in order to study similar HPC issues in symbolic computing (e.g. for the multivariate polynomial GCD algorithm).

Scientific data analysis and visualization: The research perspectives of the PEQUAN team in the field of scientific data analysis and visualization aim at addressing the upcoming challenges induced by the next generation of supercomputers (exascale computing). This includes in particular a total reboot of topological data analysis (TDA) research along 3 major directions:

— in-situ analysis: how to derive efficient and parallel TDA algorithms capable of respecting a time and/or memory budget (to allow for an easy scheduling in conjunction with the simulation) ;
— multivariate analysis: how to generalize TDA algorithms to ranges of higher dimension than one (to address the increasingly popular multi-physics simulations) ;
— uncertain data analysis: how to extend TDA approaches to uncertain data, where considered functions no longer map each point of the domain to a point on the real line but to a random variable (to address the increasingly popular parameter studies).

Next researches on motion estimation focus on a rainfall nowcasting system dedicated to the prevention of flash floods. From images, the system first computes a reliable velocity field, second makes a forecast by extrapolation. A first prototype has already been developed and daily works in a production mode. The following issues will be investigated:

— use image structures properties to improve the estimation of velocity field,
— deal with data at various scales,
— merge with punctual measures acquired by a network of pluviometers.

EQUIPE POLSYS

DOSSIER D'AUTO-ÉVALUATION

1. Présentation de l'équipe

Introduction

PolSys is a joint project-team between INRIA, CNRS and University Pierre and Marie Curie. The scientific objective of the team is to develop efficient algorithms for computing the complex solutions and/or the real ones, or the solutions in a finite field, of polynomial systems by means of exact methods (Computer Algebra). PolSys has developed several fundamental algorithms, in particular algorithms for computing Gröbner bases and algorithms based on the so-called critical point method. Among a wide range of applications that the group studies, the main focus is on applications in cryptology and, in particular, in Algebraic Cryptanalysis and post-quantum cryptography.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	4	5	4	4	4	3
Researchers	3	3	3	2	2	2
PhD students	11	10	8	8	5	4
Others (Postdoc + Emeritus + Delegation)	3	2	5	3	2	1
Total	21	20	20	17	13	10

Grants	2012	2013	2014	2015	2016	2017
International	18,04	26	30	30	30	30
National	33,73	36,41	36,38	36,38	51,48	148,33
Industry	13,94	9,8	8,25	45	45	45
Others	0	0	0	0	0	0
Total	65,71	72,21	74,63	111.38	126.48	223.33

Politique scientifique

Polynomial system solving is a fundamental problem which arises in many areas of computer science. However, a major challenge is that, in most cases, the number of solutions of a polynomial system is exponential in the number of variables. To compensate with this *intrinsic exponential behaviour*, our goal is to push forward the frontiers of current research on solving polynomial systems and let it enter in the era of high-performance computing. Our project is naturally structured through four objectives :

Objective 1 : Fundamental Algorithms and Structured Systems. Polynomial systems coming from applications are neither random nor dense, but highly structured. Various symmetries are present, but we also observe multi-homogeneity and sparsity. Even if polynomial system solving is an NP-hard problem in the worst case, structured polynomial systems are potentially much easier to solve than the general ones. This settles two fruitful research directions : design dedicated algorithms (e.g. based on Gröbner bases) taking advantage of the structure that appears in applications and identify specific configurations in polynomial systems that allows us to deduce polynomial-time algorithms; this leads to a more systematic way for modeling problems coming from applications.

Objective 2 : Solving Systems over the Reals and Applications. Solving positive dimensional systems over the reals may be very difficult because of the exponential complexity of most of algorithmic specifications. It is however very useful in applications in engineering sciences and hence there is a need of obtaining practically

fast implementations that scale enough. The key observation is that for many algorithmic problems, the best known algorithms from the practical point of view were not the ones with the best theoretical complexity. Our strategy has been to obtain algorithms whose complexity lie in the best known class but with practical performances which reflect the complexity gains. The obtained implementations have been used to solve applications that were out of reach of the state-of-the-art.

Objective 3 : Dedicated Algebraic Computation and Linear Algebra. While linear algebra is a key step in the computation of Gröbner bases, the matrices generated by the algorithms F_4/F_5 for computing Gröbner bases have specific structures (quasi block triangular). The objective is to develop a dedicated efficient multicore linear algebra package as the basis of a future open source library for computing Gröbner bases. Thus, POLSYS provides a complete suite of software for solving polynomial systems (over finite fields or real or complex numbers). This is a feature of the group.

Objective 4 : Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory. We propose to develop a systematic use of *structured systems* in so-called *algebraic cryptanalysis* (this is a general technique to study the security of cryptographic schemes). We want to improve the efficiency and to predict the theoretical complexity of such attacks. We plan to demonstrate the power of algebraic techniques in post-quantum cryptography and new areas of cryptography such as Algebraic Number Theory (typically, in curve based cryptography).

2. Produits de la recherche et activités de recherche

Bilan scientifique

Objective 1 : Fundamental Algorithms and Structured Systems. Our goal is to derive algorithms for solving polynomial systems that are efficient both from a theoretical and a practical point of view. One key way to do that is to exploit the structure and the geometry of the polynomials and of the underlying problem.

In this context we considered

Structured systems. At first glance, multi-homogeneity, weighted homogeneity (quasi-homogeneity), overdeterminedness, sparseness and symmetries seem to be unrelated structures. Indeed, until recently we have obtained specific results for one type of structure :

For bilinear systems, we obtain in [**POLSYS-C-5**] dedicated algorithm and sharp complexity results, and we can also handle symmetries in [**POLSYS-C-2**] and weighted homogeneous systems [**POLSYS-J-15**]. We also introduced sparse variants of F_4/F_5 for solving symbolically sparse systems using *sparse Gröbner bases* [**POLSYS-C-5**].

Change of Ordering. When the number of solutions is finite, a key step in polynomial system solving is the transformation of the ordering of its Gröbner basis to a lexicographical ordering. We propose algorithms [**POLSYS-J-18**] to take advantage of the sparsity. This leads to implementations which are able to handle ideals of degree over $100\,000$. We also introduced state-of-the-art complexity bounds for changing the order when the ideal is in Shape Position. Using these results we tackle algebraic problems occurring in the Discrete Logarithm Problem for elliptic curves [**POLSYS-C-7**] that were untractable before the introduction of our tools.

Complexity Analysis. Our general philosophy is to reduce the gap between worst-case and the actual behaviour of the algebraic algorithms. For generic systems, we provide a very precise bound on the number of arithmetic operations required by the F_5 algorithm to compute a Gröbner basis for the grevlex ordering which is crucial when solving polynomial systems with Gröbner bases. This is the best known result for the problem.

Fundamental Problems. We apply our novel algorithmic techniques to several fundamental problems that appear in applications, for example cryptography. Among them we mention the `Minrank` problem (the multivariate generalization of the eigenvalue problem) [**POLSYS-J-1**, **POLSYS-J-2**, **POLSYS-J-4**] and the *Isomorphism of Polynomials With One Secret (*`IP1S`*)* [**POLSYS-J-11**] where our results improve the state-of-the-art.

Guessing Recurrence Relations. We apply linear algebra techniques for finding linear recurrence relations with constant coefficients of multidimensional sequences [**POLSYS-C-11**]. This finds applications in the Gröbner basis change of ordering [**POLSYS-J-18**]. We also extend our algorithm to deal with linear recurrence relations with polynomial coefficients [**POLSYS-C-12**]. This allows us to determine whether some 3D-space walks confined in the nonnegative octant should be P-recursive or not.

Objective 2 (Solving Systems over the Reals and Applications). During the last 5 years, the main applications we focused on were the design of computer tools to automate *(i)* the stability analysis of numerical schemes for PDEs, *(ii)* the contrast problem in medical imaging and *(iii)* computational geometry (computation of Voronoi diagrams for curved objects). These were unreachable to the best computer algebra software and cross fertilization of our theoretical contributions with our implementation skills have enabled us to solve them.

Problems *(i)* and *(ii)* consist in computing some information about the projection of the real solution set to a system of polynomial equations and inequalities. They boil down to either one-block quantifier elimination over the reals [**POLSYS-J-2**] or classification issues w.r.t. parameters'values [**POLSYS-C-10**]. We provide asymptotically faster algorithms (w.r.t. the theoretical complexity state-of-the-art) which outperform the available software. Problem *(iii)* boil down to univariate solving [**POLSYS-J-3**]. We proved the optimality of the approach which requires real time computations on polynomials of significant but still moderate degrees. To go further and anticipate with the upcoming challenges, we have worked towards hybrid semi-numerical algorithms for univariate solving [**POLSYS-C-1**] (the implementation of this contribution is currently the one available in the computer algebra system MATHEMATICA) and asymptotically optimal algorithms for univariate real root refinement [**POLSYS-J-12**] while a significant software effort has been made on real root isolation (see the software SLV).

Most of the progress we made come from the theoretical study of geometric objects intrinsically related to polynomial optimization. This raised the opportunity to identify the potential impact of computer algebra in this active area. We have provided the first exact algorithm for equational constrained polynomial optimization running in time cubic in the output size [**POLSYS-J-10**] as well as bounds on the height of global minimizers which is useful to numerical procedures [**POLSYS-J-9**]. Another popular technique for polynomial optimization is semi-definite programming. We provided the first exact algorithm for solving linear matrix inequalities [**1POLSYS-J-16**] ; it has been implemented in the software SPECTRA and used to obtain a computer validation of an example contradicting a conjecture in the mathematical area named Real Algebra.

Finally, we started to work on more sophisticated algorithmic problems such as connectivity queries on the real solution sets to systems of n variate polynomial equations of degree $\leq D$. This can be used for instance to solve motion planning problems. In a series of works, we improved the longstanding complexity bound $D^{O(n^2)}$ due to Canny (1988), to $D^{O(n \log(n))}$ [**POLSYS-J-17**].

Objective 3 : Dedicated Algebraic Computation and Linear Algebra. An important part of POLSYS's activity is devoted to software development. We mention that during the evaluation we have started the development of two new open source libraries GBLA and SLV while continuing the two mature software : FGB (written in C) and RAGLIB (in Maple).

FGB. The FGB library [1] represents around 250 000 lines of code written mostly in C/C++ and Maple language. The FGB package for computing Gröbner bases is directly linked with the kernel of Maple while the Maple libraries are available as packages. Moreover, the low level software written in C (FGB) can be called *directly* from any C program ; several links with other systems (Mathemagix, Coq) have been done already. Most remarkable changes in the last evaluation periods are an experimental multi-core implementation available in Maple and a fast implementation of the SPARSE-FGLM algorithm [**POLSYS-J-18**].

GBLA. The GBLA library [2] has been jointly released with Technishe Universität Kaiserslautern (C. Eder), in January 2015. It is a new open source C library (GPLv2) for linear algebra dedicated to Gröbner bases computations, specialized in reducing matrices generated during Gröbner bases computations. The goal of the library [**POLSYS-C-8**] is to take advantage of the shape of the matrices as they are sparse and have a quasi-triangular structure. As a consequence, the new library is much faster than general libraries for this kind of matrices. A counterpart is that we offer very few functionality.

SLV. The open source C library SLV [3] has been developed in the team since 2014. It provides routines for isolating and approximating up to any desired precision the real roots of a univariate polynomial with integer coefficients. The current version of SLV [**POLSYS-J-13**] is very efficient in solving polynomials of degree 1 000 with coefficients of maximum bitsize around 10 000 bits. For the approximation step(s) SLV relies on a straightforward application of Bolzano's rule.

1. http://www-polsys.lip6.fr/~jcf/FGb/index.html
2. http://www-polsys.lip6.fr/~jcf/GBLA/index.html
3. http://www-polsys.lip6.fr/~elias/soft.html

RAGLIB. RAGLIB [4] is a free Maple package dedicated to polynomial system solving over the reals. It implements algorithms that are singly exponential in the number of variables (other software competitors implement the Cylindrical Algebraic Decomposition algorithm whose complexity is doubly exponential in the number of variables). RAGLIB is built upon FGB that computes Gröbner bases for isolating the real solutions of polynomial systems defining finite sets of points. The main functions provided by RAGLIB allow one to decide the emptiness of real solution sets to polynomial systems of equations and inequalities and/or compute sample points per connected component of their real solution set.

SPECTRA. SPECTRA [**POLSYS-J-20**] (Semidefinite Programming solved Exactly with Computational Tools of Real Algebra) is a Maple library [5] devoted to solving exactly Semi-Definite Programs. It can handle rank constraints on the solution. It is based on the FGB library for computing Gröbner bases and provides either certified numerical approximations of the solutions or exact representations of them.

Objective 4 : Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory. *Fundamental algorithms over finite fields.* A basic problem in the analysis of cryptographic primitives is to solve algebraic equations over a finite field $\mathbb{F}_q$ (and not in the algebraic closure). This is the $\texttt{PoSSo}_q$ problem. We propose a new approach, the so-called hybrid approach, that combines exhaustive search with Gröbner bases. We show in [**POLSYS-C-3**] that we can save an exponential factor in the complexity bound compared with standard Gröbner basis techniques. The special case and important Boolean case of $\mathbb{F}_2$ is considered in [**POLSYS-J-8**]. Under precise algebraic assumptions on the input system, we present an algorithm of expected complexity $O(2^{0.792 \cdot n})$ for solving quadratic Boolean systems by breaking the 2^n barrier. All in all, the set of fundamental algorithms proposed in [**POLSYS-C-3**, **POLSYS-J-8**] allow to set minimal parameters for any cryptosystem whose security is based on the hardness of $\texttt{PoSSo}_q$.

Multivariate Cryptography. This is an area of cryptography where the goal is to design cryptographic primitives based on the hardness of : the $\texttt{PoSSo}_q$ problem, the `MinRank` problem (the multivariate generalization of the eigenvalue problem) or the `IP1S` problem (given two lists of polynomials the problem is to recover a linear change of coordinates that maps one to the other). With the complexity bounds obtained on `MinRank` [**POLSYS-J-4**], we have been able to precisely asses the cost of a key-recovery attack against MQQ, HFE, Multi-HFE and variants for odd and even characteristic [**POLSYS-J-5**]. A common feature of these works is an improved modeling of the `MinRank` instance occurring in the key-recovery as well as refined complexity analysis. All the results are supported by experimental results on previously recommended parameters. In summary, we provide a state-of-the-art theoretical tools to evaluate the security (w.r.t. key-recovery) of any multivariate public-key scheme.

Cryptology on curves. Problems coming from cryptology based on algebraic curves are all very challenging. In this context, we revisit an index calculus algorithm on curves proposed independently by Gaudry and Diem in 2005. A fundamental step (the sieving one) of this algorithm relies on solving a polynomial system. We tackle this problem by considering the intrinsic structures that can help for solving these systems. In particular, we show in [**POLSYS-J-6**, **POLSYS-C-7**] how symmetries on the curve (existence of 2-torsion points) can be translated to symmetries in the polynomial systems and, in these cases, how to gain an exponential factor for the resolution of the sieving step. Efficient Gröbner basis computation [**POLSYS-J-18**] and complexity analysis of weighted homogeneous systems [**POLSYS-J-15**] were key ingredients. These recent works push some of the international experts in this domain to restart their works on index calculus on curves algorithms.

Code Based Cryptography. McEliece cryptosystem is one of the oldest unbroken public-key cryptosystem (1978). An important result of the team was to relate the security of the McEliece cryptosystem with the difficultly of solving a structured algebraic system, and more precisely, an overdetermined and multi-homogeneous algebraic system. The new results [**POLSYS-J-7**, **POLSYS-J-14**, **POLSYS-C-6**] demonstrate that algebraic cryptanalysis is an important technique to asses the security of codes-based cryptography. In [**POLSYS-J-7**], we design the first method allowing us to distinguish alternant and Goppa codes over any field provided that the codes have sufficiently large rates. It was widely believed that this problem was a hard decision problem and was an essential component of the security proof of McEliece. In [**POLSYS-J-14**], we show that all compact variants of McEliece proposed in the literature have a fundamental weakness. The security of such compact variants does not rely on the bigger compact public matrix but on a scaled down McEliece scheme without any symmetry. We believe that this result challenges the core principle of using compact variants of McEliece.

4. `http://www-polsys.lip6.fr/~safey/RAGLib`
5. `http://homepages.laas.fr/henrion/software/spectra/`

Algebraic Methods Applied in Side Channel Attacks. Side Channel Attacks are any attacks based on information gained from the physical implementation of a cryptosystem. From an algebraic cryptanalysis point of view it means that we have additional equations that can be used to improve the solving of the underlying algebraic problem. We applied this approach in the context of asymmetrical cryptography to attack RSA. Coppersmith introduced in 1995 an algebraic method, based on lattice reduction, to factor an RSA integer (product of two distinct large primes) when some part of one of its factors is known. During the evaluation period, we proposed the first theoretical and practical improvements over the Coppersmith's seminal algorithm. From a practical point of view, problems taking years of computation can now be solved within some days. Since this algorithm is the keystone of a lot of attacks on RSA, this result has some impact in Cryptology but our approach can be more generally used in LLL reduction.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	12	18	4	4	11	6	55
Ouvrages et édition	3		1	1	1		6
Conferences	18	7	12	10	14	3	64
Logiciels	1	1		1	2	1	6
Brevets				1			1

Faits marquants

— In [**POLSYS-C-4**], we improve the complexity of Index Calculus Algorithms in Elliptic Curves by means of Gröbner basis techniques and we analyze the complexity of this new approach by using the multi-homogeneous structure of the equations.

— In [**POLSYS-C-1**], we present an efficient algorithm to isolate the real roots of a univariate polynomial the coefficients of which are polynomial functions of real algebraic numbers. An implementation of this algorithm in C is now part of the core library of MATHEMATICA.

— In [**POLSYS-J-17**], we design the first nearly optimal algorithm for answering connectivity queries in the real solution set of polynomial equations. This improves the long-standing upper bound from J. Canny (1988). We expect from these results the first practical algorithms and software for solving this algorithmic problem which finds applications in several engineering sciences.

— Berlekamp-Massey is well known algorithm with many applications in Coding Theory, Computer Algebra. We generalized to the multidimensional case [**POLSYS-C-11**, **POLSYS-C-12**].

— Jointly with Univ. Of Kaiserslautern (C. Eder), we have released a new open source C library for linear algebra – GBLA – dedicated to Gröbner bases computations. This new library opens the door to high performance applications.

— We initiate a collaboration with B. Sturmfels (University of Berkeley) to study Geometry and Optimization problems with ALgebraic methods. This result is published in Journal of algebraic statistics. Another outcome of this collaboration is the joint project with MIT underlying the Marie Curie grant of Kaie Kubjas who will come for one year in our group in 2018.

— In Dec. 2016, the RISC project was one of few projects accepted in the "Grands Défis du Numérique" program (PIA). The goal of the RISQ project is to prepare the security industry to the upcoming shift of classical cryptography to quantum-safe cryptography. The RISQ project is a massive effort at the French level to embrace the quantum-safe revolution. The project gather 15 academic and industrial partners. The RISQ project is certainly the biggest (in term of number of partners, as well as funding) industrial project ever organized in quantum-safe cryptography.

— Jointly with LAAS (D. Henrion, S. Naldi), we have released a new MAPLE library SPECTRA for finding a real point $x = (x_1, \ldots, x_n)$ such that the symmetric matrix $A(x) = A_0 + A_1\, x_1 + \cdots + A_n\, x_n$ is positive semidefinite using exact arithmetic.

— Our open source C library SLV has been officially released this year with a presentation at ISSAC. It aims at isolating and approximating the real roots of univariate polynomials with integer coefficients.

— Matías Bender received the Distinguished Student Author Award of ISSAC2016 for his paper written with J.-Ch. FAUGÈRE, L. PERRET and E. TSIGARIDAS.

3. Analyse SWOT

Strengths. The team has a strong expertise ranging from theory (algebraic geometry, algebraic complexity theory) to applications (statistics, control theory, ...), industrial applications in cryptology (with Oberthur, Gemalto, ...) and mechanics (with Dassault); Applications are used to define our theoretical priorities and to evaluate the efficiency of our algorithms and implementations.

The group has also done pioneering work in Algebraic Cryptanalysis which reduces the security analysis of a cryptographic primitive to the difficulty of solving a polynomial system. This technique has become a standard one in cryptography. This expertise is at the origin of our industrial transfer activity.

We have started the development of two new open source libraries GBLA and SLV while continuing the two mature software : FGB (written in C) and RAGLIB (in Maple), providing a complete suite of software for solving over finite fields or real or complex numbers. This gives us strong visibility in some application domains. A witness of the international impact of the team is that most of commercial or academic computer algebra systems provide implementations of algorithms designed by the group (e.g. for Gröbner bases).

The team has also developed strong international collaborations in Europe and Canada, China and the USA. It is the French component of the Inria associated team GOAL with UC Berkeley (B. Sturmfels) whose aim is to develop algorithms and mathematical tools to solve geometric and optimization problems through algebraic techniques.

Weaknesses. The size of the group can be problematic in the future, in particular to maintain a wide activity in algebraic cryptanalysis.

Also, the nature of our research topic, at the interface of algebraic geometry and computer science, makes difficult to find students appropriately trained to prepare a PhD. The current development of joint curricula between mathematics and computer science in which we are highly involved at UPMC may help.

4. Projet scientifique à cinq ans

Objective 1 : Fundamental Algorithms and Structured Systems. SPARSE POLYNOMIALS. In [**POLSYS-C-5**] we have proposed variants of F_5 and $FGLM$ to solve symbolically systems whose support lie in the same support (unmixed systems). We want to investigate two new research directions. First, we will study the generalization to mixed systems (where the algorithms and the complexity would depend on the Newton polytope of *each* of the polynomials of the system). The other direction would be to study the case where only a small subset of monomials appear in the equations (*fewnomial systems*). We explore the complexity of computing a sparse Gröbner basis when the number of monomials in the support is close to the number of variables.

From an implementation point of view, we believe that implementing efficiently this new generation of algorithms in a general framework is a new research area. Algorithms from convex geometry or the combinatorial world would be necessary components of an efficient implementation.

SPARSE-FGLM/BERLEKAMP – MASSEY – SAKATA. Given a n-dimensional linear recursive sequence, the so-called BMS algorithm computes a Gröbner basis of the ideal of relations. As the sparse/fast FGLM [**POLSYS-J-18**] is related to the efficiency of BMS, it seems natural to investigate potential improvements of this algorithm. In particular, as the output of BMS is a Gröbner basis of the ideal relation, we aim at investigating to what extent we can adapt recent developments on Gröbner basis algorithms to BMS. A byproduct would be new complexity results for the FGLM algorithm depending on the geometry of the staircase.

ISOMORPHISM OF POLYNOMIALS. As our algorithm [**POLSYS-J-11**] solves the `IP1S` for quadratic polynomials verifying a regularity assumption, we plan to extend our algorithm to all instances, *i.e.* the so-called *irregular case* as well. We also aim to solve `IP1S` for polynomials of higher degrees. For degree-3 polynomials, this problem is deeply connected to Graph Isomorphism, the latter reduces to the former. Therefore, this more general problem is intrinsically difficult. However, we aim at determining subclasses of polynomials sets such that `IP1S` can be solved in polynomial time. A more general isomorphism of polynomials `IP` also exists, with two secrets. We have already developed efficient algorithms for instances of these problems used in cryptography. Yet, the algorithms proposed so far were usually relying on heuristic assumptions. We would like to find families of IP instances for which we can prove polynomial-time complexity.

FUNCTIONAL DECOMPOSITION PROBLEM (`FDP`). This fundamental problem in computer algebra finds application in cryptography. Given a set of homogeneous polynomials, we should compute two sets of homogeneous polynomials such that the former is the composition of the latter two. Although the input is not the same,

this problem is linked to the IP1S. As this problem is somewhat too general, a dedicated algorithm could always return a trivial answer. Therefore, we plan to add constraints on either the inner or the outer polynomials, so that the output of the designed algorithm, if possible, is of some interest.

Guessing Recurrence Relations. Both our algorithm [**POLSYS-C-11**] and the Berlekamp – Massey – Sakata algorithm extend the Berlekamp – Massey algorithm to multidimensional sequences in different ways. We aim at studying in-depth their resemblances and their differences. A by-product of this study would be the design of an hybrid algorithm taking advantage of the strengths of both algorithm with new complexity bounds.

Objective 2 (Solving Systems over the Reals and Applications). Our primary goal remains unchanged : obtain practically fast implementations. Our methodology is to reduce the gap between optimal theoretical complexity and practical efficiency. We will continue the line of work already started but we will also investigate three research directions that would allow to push forward the current limitations of polynomial system solvers and intensify our efforts on the application side.

The first one consists in designing dedicated algorithms that exploit some structures on the input systems to be solved (sparsity, symetries, multi-homogneity, etc.). many progress have been made on the algebraic phase (Gröbner bases) in the zero-dimensional case but much remains to be done in algorithms for solving over the reals. We expect quantitative results (e.g. on the decrease of the output size induced by some structures) and better algorithms.

The second one is to go further on algorithms for deciding connectivity queries and take advantage of the complexity breakthrough which has been obtained. Theoretical developments are still needed for handling the case of systems of equations *and* inequalities. As importantly, we want to start the development of a software package for solving such algorithmic problems and tackle challenging applications coming from robotics. We also intend to focus on other algorithmic problems that help to provide informations on the *topology* of solutions sets over the reals to polynomial systems of equations and inequalities.

Last, but not least, we need to improve the univariate solving step significantly. It is a challenge to close or to reduce the current gap between the best theoretical complexity bound and the subdivision algorithms used in practice. A first step in this direction is to exploit the observation that even when the degree and the coefficients of the input polynomial are huge, most often, the number of bits that we need to isolate the real roots is not that big. This, in conjunction with semi-numerical subdivision techniques, will lead to an improved algorithm whose substantial merit is its simplicity, important for the implementation. There are current works on Descartes and we plan to apply similar techniques to CF.

Objective 3 : Dedicated Algebraic Computation and Linear Algebra.

FGB, GBLA The most efficient algorithms (e.g. F_5) rely heavily on reductions to row echelon forms (linear algebra) on structured matrices. We aim at enhancing these computations by improving the open source linear algebra library GBLA [**GBLA2016**]. One direction is to improve the parallel scaling on multi-core CPUs and also take advantage of GPUs using interfaces like OpenCL.

The other main part of the whole algorithm is to generate the matrices. To avoid a possible bottleneck, this part also needs new, parallel implementations. The crucial parts that need to be implemented are an efficient hashing for representing the monomials, the so-called *symbolic preprocessing* that searches for possible polynomial reducers and the construction of the matrices for GBLA. There is here a risk factor : in general, the problem of defining parallel friendly data structures that can be efficiently created, updated, and accessed is a significant research challenge. Hence, all these steps need to be parallelized carefully.

SLV The goal is to provide an open-source, efficient and reliable software for solving univariate polynomials over the reals, either as a stand-alone package or integrated to well-known computer algebra systems that can handle huge polynomials and cluster of roots. We will integrate it with the best Gröbner basis engine, FGB, to provide a state-of-the-art tool for solving exactly 0-dimensional polynomial systems. Theoretical advances by modifying existing algorithms are also in place.

RAGLIB For the next years, the first task is to take advantage on the recent progress made on Gröbner bases computations and available in FGB. That may lead to several important practical improvements. We also intend to make our future implementations of algorithms answering connectivity queries available in RAGLIB as well as our new algorithms for performing quantifier elimination over the reals.

SPECTRA The goal for SPECTRA is now to integrate reduction tools which allow us to reduce the size of the input data by analyzing the structure of the input linear matrix inequality. This will be very useful to compute certificates of positivity for multivariate polynomials. Another goal is to give the ability to SPECTRA to study the intersection of several solution sets to Linear Matrix Inequalities. This will be useful to compute certificates of positivity of polynomials up to some constraints.

Objective 4 : Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory. In this objective, we focus on solving $\texttt{PoSSo}_q$ over finite fields and the corresponding applications (Cryptology, Error correcting codes,...). Our primary goal remains to make progress on the security analysis of challenging hard problems in public-key cryptography. The methodology is to improve algorithms solving fundamental problems such as $\texttt{PoSSo}_q$, `MinRank`, `IP1S`, finding the small roots of a polynomial system, and to use structured systems to improve toward existing techniques. A new goal of this objective is to make industrial transfer in multivariate cryptography.

Sparse Gröbner Bases in Algebraic Cryptanalysis. It is often the case that the algebraic systems occurring in algebraic cryptanalysis are *structured* (for example, in multivariate or code-based cryptography). We would like to investigate the new possibilities offered by sparse Gröbner Bases [**POLSYS-C-5**] in algebraic cryptanalysis. An objective is to investigate whether this feature can be exploited by the complexity analysis done in [**POLSYS-C-5**]. This research direction will be conducted in close interaction with **Objective 1**.

Mastering the Theoretical Complexity. A challenging complexity objective is on the genericity assumption used in the analysis of the complexity of the Arora-Ge Algorithm against LWE. The problem is to prove that dth powers of m generic linear forms in $n \geq m$ variables forms a semi-regular system. This is a fundamental question essentially equivalent to the well-known Fröberg's conjecture which is considered as difficult. In the next evaluation period, we would like to investigate the conjecture in some restricted settings.

Applications to Algorithmic Number Theory. Encouraged by the first results we obtained during the last period of evaluation, we consider to continue our research in this domain. One of our main result is the new variant of Coppersmith's algorithm for finding small roots of univariate modular equations. We plan to tackle the generalization of this problem to polynomial systems. This is a challenging problem and has a wide spectrum of potential applications in cryptology. First, we want to analyze the connection between Gröbner basis computations and lattice-based methods developed for solving instances of this problem. Our main objective is to find if some structures appear during the solving step and try to use them. In the context of curve-based cryptology, we mainly plan to implement an efficient version of the dedicated algorithms we have designed during the last period of evaluation and confront them to real size discrete logarithm problem over elliptic curves.

Technology Transfer in Multivariate Cryptography. Two of the most important standardization bodies in the world, NIST and ETSI have recently started initiatives for developing cryptographic standards not based on number theory, with a particular focus on primitives resistant to quantum algorithms. An objective here is then to focus on the design of multivariate schemes. The team holds a solid expertise in the analysis of such quantum-resistant schemes. Besides the development of new cryptosystems, this includes proposals for quantum-resistant standards to the NIST competition as well as transfer to an industrial partner. We emphasize that this research/innovation direction will take benefit of our involvement in the PIA RISQ.

DEPARTEMENT **DAPA**

1. Présentation du département

The DAPA *Data and Knowledge Science* research department has been created in 2006. DAPA has been conducted by Bernadette Bouchon-Meunier from its creation and Bernd Amann is at the head since her departure in 2014. The department has currently 24 permanent members and about 60 PhDs and post-doc students and (temporary) engineers :

Workforces	June 30th, 2017
Professors PR/MCF	7/15
Researchers DR/CR	0/2
PhD	48
Others*	13
Total	85

* : engineers, ATER, post-docs, invited researchers (> 6 months)

It is composed of four research teams :

ACASA : Cognitive Agents and Symbolic Machine Learning
BD : Databases
LFI : Learning, Fuzzy and Intelligent systems
MLIA : Machine Learning and Information Access

2. Politique scientifique

Scientific activities and goals : DAPA's research activities cover a wide spectrum of methods from Data Science, Artificial Intelligence, Digital Humanities and Big Data :

— Information retrieval in text, images, multimedia and streams (MLIA, LFI, BD)
— Knowledge representation and reasoning (LFI, ACASA)
— Heterogeneous, distributed and dynamic databases (BD)
— Cognitive sciences and digital humanities (ACASA)
— Recommendation, collaborative filtering and information personalization (MLIA, LFI, ACASA, BD)

These methods are applied for dealing with advanced data and knowledge-centric problems generated by social media and other web applications, in the IoT domain, in human sciences, etc. DAPA's expertise is valorized through many national and international research contracts (EU, ANR, Labex) and industrial collaborations (FUI, CIFRE, CLEAR joint lab with Thales) for financing students and engineers. During the evaluation period, the DAPA department has raised fundings of more than 10ME.

At the University level, DAPA is strongly involved in the two UPMC Institutes ICSD and IUIS and participates in the two Labex programs SMART and Obvil. DAPA members also actively participate in various committees at the different administrative levels (university, faculty, teaching departments, laboratory) [1].

At the national level, two DAPA members (A. Doucet, M. Cord) are engaged as scientific delegate at the INS2I CNRS institute and one member (J.G. Ganascia) chairs the CNRS Ethical Committee. During the evaluation period, two DAPA members (M. Cord, J.G. Ganascia) have been nominated to the Institut Universitaire de France (IUF) .

1. See team presentations for more details

At the international level, DAPA collaborates with other teams on all continents and DAPA research members are also regularly invited in program committees of major conferences and journals and other scientific evaluation committees (PhD, grants). During the evaluation period, DAPA also has organized international conferences and competitions and obtained best paper awards and research grants [2].

DAPA's research expertise and is also a major driving force in several UPMC Master programs (Erasmus Mundus Master Program DMKM ; Master DAC in data and knowledge science ; Big Data Certificate).

In conclusion, DAPA covers a coherent range of scientific activities which perfectly respond to the current demand in science and industry. Data-centric artificial intelligence has been identified as a major challenge over the next 10 years (cf. National Strategy in Artificial Intelligence "FranceIA") and DAPA aims at enforcing its presence and contributions in these domains at the national and international level.

Animation : Each research team has its specific research domains is independent in the definition of its fundamental scientific goals. The main role of the department consists in the more general scientific coordination and animation :

- Scientific Seminar : During the evaluation period, the department has organized 65 seminars of invited external researchers `http://www-dapa.lip6.fr/web/?q=seminaires` in complement to the LIP6 Colloquium.
- Scientific Events : DAPA also regularly organizes social and scientific events (once or twice a year) complementary to the internal team meetings and the LIP6 PhD seminar.
- Employment profiles : The different teams regularly discuss about their evolution for the preparation of permanent employment profiles.
- Education : All DAPA teams are strongly involved in the Data and Knowledge Science Master program (DAC) which gives the occasion to discuss together about current trends and needs with respect to the specific research domains.
- Visibility : Through its wide but coherent research expertise, DAPA is visible as a department and serves as entry point for establishing a first contact for industrial partners with specific scientific questions and contract opportunities.

2. See team presentations for more details

EQUIPE ACASA

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

Originally, the research interests of the ACASA team were focused on Artificial Intelligence, Symbolic Machine Learning, Data Mining, Knowledge Acquisition, Scientific Discovery and Creativity. With time, they have evolved and are now focused on cognitive modeling, especially on modeling Ethical Systems, Social Representation and Argumentation, and on studying the literary side of Digital Humanities, in partnership with the literary teams of the Paris Sorbonne University, within the OBVIL Labex of which the ACASA team is one the pillars. In addition, we pursue an activity on Symbolic Data Fusion in collaboration with the LIP6-SOC team. The meaning the of ACASA acronym reflects this thematic evolution : while originally it signified "Acquisition des Connaissances et Apprentissage Symbolique Automatique" today it denotes "Agents Cognitifs et Apprentissage Symbolique Automatique".

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	4	3	3	3	4	4
Researchers	1	4	7	7	7	3
PhD students	4	6	6	6	6	4
Others		1		1	1	2
Total	9	14	17	17	17	13

Remark : it has to be noted that some of the PhD students are not recorded as LIP6 members when they are registered with other institutions, in particular with the OBVIL Labex, but when they are working jointly with the ACASA team, under the supervision of a member of the ACASA team.

Grants	2012	2013	2014	2015	2016	2017
International	0 k€	0 k€	0 k€	0 k€	42 k€	85 k€
National	121 k€	196 k€	230 k€	103 k€	117 k€	79 k€
Industry	18 k€	9 k€	3 k€	0 k€	0 k€	0 k€
Others	0 k€	0 k€	0 k€	3 k€	15 k€	15 k€
Total	139 k€	205 k€	233 k€	106 k€	174 k€	179 k€

Remark : the ACASA team benefits of some scholarships for PhD or post-doctoral students from the OBVIL Labex of which it is an active member. However, the corresponding resources don't appear in this table.

Politique scientifique

The activity of the ACASA team focuses on two main axes : cognitive modeling and digital humanities. On the cognitive modeling side, the ACASA team is both working on the modeling of Ethical Systems and on the modeling of social representations based on psychological and sociological theories.

The modeling of Ethical Systems began 10 years ago with the works of Jean-Gabriel Ganascia on the use of Answer Set Programming to simulate a few different ethical systems present in the philosophical tradition. Now, it takes place within the ANR Ethicaa project, which involves different research teams on the ethics for artificial agents. A PhD student, Fiona Berreby, is working under the supervision of Gauvain Bourgne and Jean-Gabriel Ganascia on the use of action language to model different kind of ethical theories. She proposed

a modular framework implemented in ASP to model multiple ethical reasoning, in particular the Doctrine of Double Effect, which is a traditional philosophical approach that aim solve some classical ethical dilemmas. Concerning action language, there is also an ongoing collaboration with LIPN on collaborative learning of action theories and models of collective learning are investigated by a PhD student, Lise-Marie Veillon, under the supervision of Henry Soldano, from LIPN, and Gauvain Bourgne. Gauvain Bourgne has also explored collective deductive reasoning with Katsumi Inoue and Nicolas Maudet by working on multiagent consequence finding.

Last year, Thomas Powers (Professor, University of Delaware, USA) spent 4 months in the ACASA team. The ACASA team is also leading the ethical package in the Emospace ITEA3 project. Lastly, the ACASA team has been involved in many international workshops, conferences and books on that topic, for instance in ECAI, AAAI or with the Japanese teams of Tokyo (InoueLab, National Institute of Informatics) and University of Yamanashi. Jean-Gabriel Ganascia has also been invited to give talks and to participate to a bilateral French-Japanese symposium on Artificial Intelligence and Ethics.

The second aspect of the cognitive modeling activities of the ACASA team began more than ten years ago with the extraction of social stereotypes from newspaper articles. When Julien Velcin, who was very active in this field, left the ACASA team to get a teaching position in Lyon, the thematic wasn't able to maintain its activity. However, today, it seems to be possible to reactivate this thematic for three reasons : the arrival of Colette Faucher in the ACASA team, the obtainment of an HDR by Julien Velcin, who would like now to collaborate with us and the current digitization of press, both of past 19^{th} and 20^{th} century press by the BNF ("Bibliothèque Nationale de France" that means *National French Library*) and of contemporary press, because of the online presence of almost all the present newspapers. More precisely, Colette Faucher is already working on a cognitive modeling approach based on theories issued from social psychology, sociology and psychology of emotions. She is starting a new project funded by the DGA ("Direction Générale de l'armement" that is in French *General directorate of the armament*) in collaboration with the LPS, a laboratory of social psychology from Aix-en-Provence, and the LEM, a laboratory of the EPHE ("École Pratique des Hautes Études"). This project aims to fight the religious radicalization of fanatic Muslims by developing different software aiming to inform the general public about the real risks of radicalization and to prevent people, who are near to be radicalized, by using a cognitive-based strategy.

The activity on the literary side of Digital Humanities lies in the prolongation of our prior works on text mining, text classification, uni-lingual alignment, name entity recognition and extraction of musical and textual recurrent patterns. This activity takes mainly place within the OBVIL Labex, of which Jean-Gabriel Ganascia is Deputy Director. Many of the resources (PhD students and post-docs) of the ACASA team come from this Labex, which is a joint action of the ACASA-LIP6 teams of UPMC and the literary teams of Paris-Sorbonne University funded by the French Government. There are many national and international collaborations in this context, for instance the "Use and Reuse" project that is funded by the Mellon Foundation, in which the ACASA team collaborate with the *ARTFL project* of the University of Chicago, the ANR Phœbus project or the presence for two years of Glen Roe, who is funded by the prestigious Australian Research Council. Besides, our scientific policy is to develop a sustainable activity in developing and maintaining tools for the digital humanity scientific community in collaboration with different partners, in particular with the Paris-Sorbonne university and with the BNF with which we are now actively collaborating through the OBVIL Labex. Lastly, the ACASA team is involved the the Prada DIM ("Domaine d'Intérêt Majeur" that involves funding from the Île-de-France region) that is intended to promote research on the literary side digital humanities at a regional scale.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Besides the publication in international journals (i.e. DH journals), in international conferences of our domain of activities (AMAS, AAAI, etc.), and in books, we are now an important actor both in the field of Digital Humanities. The existence of the OBVIL Labex, the DIM *Prada* or the nomination of Jean-Gabriel Ganascia as senior member of the IUF ("Institut Universitaire de France") testify of this recognition. Funded by the Mellon Transnational program, our international collaboration with the ARTFL project of the University of Chicago is also a sign of reconnaissance. Lastly, the number of software (e.g. Medite `http://obvil.lip6.fr/medite/`, Phœbus `http://obvil.lip6.fr/phoebus`, DeSeRT `http://obvil.lip6.fr/`

desert, EReMoS http://eremos.lip6.fr, Comics++ http://comicsplus.lip6.fr, Dissimilitudes http://dissimilitudes.lip6.fr:8180/#/ and REDEN) that we are developing and maintaining for the Digital Humanity community are also significant results of our scientific activities.

Our activities in the field of computational ethics deserves also to be mentioned, because it begins to be recognized as a valuable contribution in the scientific community. The active participation to the ANR EthicAA project as well as the publications in highly ranked international conferences, as is AMAS, attest to this.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	1	0	2	4	2	6	15
Ouvrages	2	0	3	2	0	3	11
Conferences	2	6	5	9	7	4	33
Logiciels	1	0	1	2	4	8	8
Brevets	0	1	0	0	0	0	1

Sélection des produits et des activités de recherche

Faits marquants

- **OBVIL Labex** : the ACASA team is one of the pillars of the OBVIL Labex. It encourage the collaboration of the computer scientists of the UPMC with the literary teams of Paris-Sorbonne University to work on the literary side of Digital Humanities. It has been a great success to have this interdisciplinary Labex selected. During these last four years, the many technical achievements we made and our numerous publications that have been accepted have shown that this has also been a success from a scientific point of view.
- **DIM** *Prada* : the ACASA team is involved in the Prada DIM that is intended to promote research on the literary side of digital humanities in the Île-de-France region. This offers us a very challenging opportunity to pursue fruitfully our researches.
- **MEDITE, Phœbus, DeSeRT and EReMoS Software** : theses different software, which result from our researches, have been developed and maintained by the ACASA team for the needs of the digital humanists. They are freely available online.
- **"Use and Reuse"** : funded by the Mellon foundation, this project offers a unique opportunity to engage a transnational collaboration with one of the most reknown teams in the world in the field of digital humanity. Our goal is to develop a joint software that aggregates our algorithms to automatically detect reuses, citations and borrowings on huge quantities of texts (130,000 books) that have been given by the BNF.
- *Best Paper Award* : Marine Riguet and Suzanne Mpouli whore are PhD students in ACASA team got the best paper award at the Digital Humanities conference 2016 in Krakow (Poland).
- Nomination of Jean-Gabriel Ganascia as **senior member of the IUF** ("Institut Universitaire de France")
- Nomination of Jean-Gabriel as the **chairman of the COMETS** that is the CNRS Ethical Committee.

3. Analyse SWOT

- *Strengths* : The main forces of our research team come from the involvement of many national and international projects, in particular the OBVIL Labex, the DIM (Domaine d'intérêt majeur) *Prada*, the "Use and Reuse", the LOCUPLETO, the EMOSPACE, the ECRICOL, the Phœbus and the EthicAA projects. In addition, Colette Faucher who is arriving in the ACASA team, have a project with the DGA ("General Directorate for the Armament") that can contribute to our research activities.
- *Weaknesses* : Our main weakness comes from the very low number of permanent researchers. The arrival of the Colette Faucher will partially remedy this problem, but only partially. One solution would be to recruit a young researcher in the field of Digital Humanities. In addition, we would like to be more present on the international scene. Even if we have a few papers accepted in renowned scientific journals and conferences, it would be suitable to increase our leadership in the scientific community.

— *Opportunities* : The new Sorbonne University constitutes a great opportunity to develop our partnerships with the literary teams of the Paris-Sorbonne university and with the BNF to constitute a world center in digital humanities. The involvement in the *Prada* DIM, the BNF repository of 130,000 books of which we have obtained the usufruct and the collaboration with the ARTFL project of the University of Chicago are of a great help to reach this goal. The arrival of Colette Faucher is an opportunity to develop our cognitive modeling activities. The nomination of Jean-Gabriel Ganascia as chairman of the COMETS and of member of the CERNA reinforces our leading role not only in the french computer ethics community, but also in the field of computational ethics.

— *Threats* : The different tools we would like to develop and to make available to the scientific community require a lot of engineering. To achieve the role we would like to play, we need one or two good permanent research engineers that could develop and maintain our tools. This is a critical point for our future. Until now, we have only employed post-doctoral researchers for this task, but it has clearly appeared that is was not sufficient.

4. Projet scientifique à cinq ans

One the one hand, we would like to pursue modeling of Ethical System and the investigation of the ethics of autonomous systems, which looks to be very important today with the proliferation of autonomous vehicles and more generally robots. Number of national and international collaborations on this topic are now under investigation.

On the other hand, we want to pursue our collaboration with the literary teams of the Sorbonne within the framework of the OBVIL Labex. Our aim is to develop and to validate many new operators of interpretation that make use of artificial intelligence techniques. There are already a few tools available online (e.g. Medite `http://obvil.lip6.fr/medite/`, Phœbus `http://obvil.lip6.fr/phoebus`, DeSeRT `http://obvil.lip6.fr/desert`, EReMoS `http://eremos.lip6.fr` and REDEN) that are currently used by literary researchers and that we would like to test on big corpuses and to distribute to the international community of scholars. We plan to develop some of these tools in collaboration with other universities. This is in particular the case with Medite and Phœbus that we plan to merge with Philoline currently developped by the ARTFL project of the University of Chicago. This could also be the case with the syntactical pattern extraction techniques that we plan to develop with a few international universities.

Lastly, an official agreement between the OBVIL Labex and the French National Library (BNF – Bibliothèque National de France) will give us the possibility to exploit 130000 digitized books in French, which corresponds to the biggest digital library in French. Our long term goal, after the end of the OBVL Labex, would be to have an UMR (Unité mixte de recherche) between the Sorbonne University COMUE, the BNF and possibly the CNRS, which would enrich and make available all these texts and all the tools we are developping to the community of Scholars throughout the world.

ÉQUIPE BD

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The BD (Bases de Données) database research team has been created by A. Doucet in 1994 and is since 2006 part of the DAPA (Données er APprentissage Artificielle) department. B. Amann is the team leader since 2014. From its creation until now, the principal research topics of the BD team concern data transaction and query processing in distributed environments. Within this context our goal is to propose optimal and scalable data processing solutions by using data distribution and replication, indexing and other physical, logical and/or semantic optimization techniques. Our research contributions mainly focuses on the second LIP6 research axis "Data science, intelligence and optimization".

Tableau des effectifs et moyens de l'équipe

The team is composed of 2 full professors (PR), 4 assistant professors (MCF) and one CNRS researcher (CR). One assistant professor has the "Habilitation à Dirigier des Recherches" (HDR), one member is currently preparing his HDR and a third member is a potential candidate during the next evaluation period. One team member is half-time engaged as scientific delegate at the INS2I CNRS Institute. From January 2015 to December 2016, the team hosted one associate researcher from Fraunhofer and two research engineers in the context of the two-year Fraunhofer-SU project EBITA. Eight students have successfully completed their PhDs during the evaluation period.

Workforces	2012	2013	2014	2015	2016	2017
Professors	6	6	6	6	6	6
Researchers	1	1	1	1	1	1
PhD students	8	8	7	7	4	4
Others*	0	0	3	5	6	5
Total	15	15	17	17	17	16

* : engineers, post-docs, invited researchers (> 6 mois)

During the last four years the team has participated in seven industrial and scientific research grants for a total amount of about 828 kE used for financing the team activities and PhD students. The total budget (in kE) does not include national or foreign public PhD scholarships (e.g. allocations MENRT). CIFRE contracts appear as industrial contracts.

Grants (kE)	2012	2013	2014	2015	2016	2017	**total**
International	28.79	28.71	14.24	0.00	0.00	0.00	**71.73**
National	0.00	0.00	236.95	245.79	166.00	79.85	**729.91**
Industry	19.73	22.25	16.06	37.91	50.45	37.63	**184.03**
Total	48.51	50.96	267.24	283.70	216.45	115.81	**982.69**

Politique scientifique

The main research goal of the BD team is to propose efficient solutions for implementing complex information processing tasks. Our research approach is built on top of the traditional data abstraction principle separating application logics (programs) from a generic data processing layer. Our solutions generally build

on top of existing customized data processing layers for specific complex and voluminous data (graphs, sequences, streams, semi-structured) and achieve efficiency by developing new algorithms on top of advanced data processing technologies including data distribution and replication, query optimization, indexing and data parallelism. Our research is driven by various application domains (data science, web archives, social media, web advertising) and the related research contracts (ANR, CIFRE, Europe, FUI) are mainly used for hiring students who are the backbone of our research. Another important motivation is to establish national and international scientific collaborations and to collaborate with industrial partners for understanding current important practical challenges and validating our research solutions.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Data streams and continuous queries : (B. Amann) The web produces continuous streams of text items published as RSS news, tweets, blog messages etc. Users can subscribe to these streams be defining queries which continuously filter and rank the most recent information items. A major challenge is then to efficiently process millions of such subscription queries over high rate input streams. In the context of the ROSES ANR project ROSES (2008-2012) on RSS feed aggregation and filtering, we worked in on multi-query optimisation (PhD J. Creus [BD-T-7]), on efficient refresh strategies for dynamic RSS feeds (PhD of R. Horincar in collaboration with the MLIA team [BD-T-8,BD-MLIA-J1,BD-MLIA-C-2]), and on continuous top-k query processing (PhD of N. Vouzoukidou in collaboration with ICS-Forth, Crete [BD-T-5,BD-C-15,BD-D-5]).

Web archive indexing and maintenance : (S. Gançarski, B. Piwowarski, A. Doucet) The Web is continuously changing and building representative archives for the future generates many challenging data processing problems including efficient harvesting, storage and indexing of Web resources. Our work on these challenges has started within the ANR Cartec project (before 2012) and pursued in the context of the European research project SCAPE (2011-2014) and an ongoing collaboration with the Central University of Venezuela (UCV), Caracas. The originality of our contributions is based on a Web page segmentation and ranking model which identifies and ranks semantically coherent blocks within a web page. Based on this model, highly-ranked blocks can be considered as more important by a page refresh strategy or more relevant within an information retrieval task. Three PhD thesis took place in this context on (1) semantic web page refresh strategies [BD-J-3] for web archives, (2) Web archive querying and text indexing (pruning) [BD-T-6], and (3) Web page segmentation and migration to HTML5 format [BD-T-4,BD-C-8,BD-C-11] (collaboration with UCV). A collaboration with the MLIA team (PhD thesis of M. Law [BD-T-3,BD-MLIA-C-3]) allowed us to explore advanced machine learning and image analysis techniques to improve the accuracy of our page change estimation model.

Workload-aware data replication : (H. Naacke, S. Gançarski) Distributed transactions in large data clusters generate a high control and synchronization overhead which is a major obstacle for achieving scalability. To reduce this overhead, we focus on user-centric/social applications where (1) the data fragment attached to each user defines the basic access unit, (2) transactions mostly access the data of two users (message exchange) and (3) the access frequency (popularity) is biased and fluctuates over time [BD-C-14]. To achieve optimal performance, we propose to move user data to a single node where the transaction can be executed locally. Then, under the assumption that users interact within social circles, we detect data groups (or bundles) and adapt data placement and cluster resources gradually according to their interactions. We have studied a related problem in the context of the GBIF (Global Biodiversity Information Facility) community and designed advanced querying facilities for world-wide federations of autonomous databases storing descriptions of observed natural species. To overcome the limited query capacities of GBIF database interfaces, we proposed a distributed middle-ware providing a higher-level structured access to the GBIF data. Our main contributions are a decentralized algorithm for parallel query processing and a cost-based placement and replication strategy keeping node usage (storage and query processing) below a given upper bound. This work is part of two PhD thesis (I. Gueye and N. Bame [BD-T-1]) in collaboration with the University Cheikh Anta Diop (UCAD) in Senegal.

SPARQL query optimization : (H. Naacke, B. Amann, M.A. Baazizi) RDF has become a standard format for publishing and integrating data and knowledge on the web. The resulting Linked-Open-Data cloud is conti-

nuously growing and implementing efficient SPARQL query processors is a challenging problem. To achieve scalability, we propose to exploit MapReduce data parallelism implemented in the Apache Spark platform. For this, we defined a cost-model which allows to choose among two standard distributed join operators (partitioned and broadcast join) implemented on top of the different Spark data layers. Our experimental results show that hybrid join plans combining these two operators allow to improve the performance significantly [BD-C-3].

Social Recommendation Models and Algorithms : (C. Constantin, H. Naacke, B. Piwowarski) Recommendation methods aim to predict, for a given user, the rating or preference she would give to an "item" which can be an action, a document, a product or even another user etc. These predictions can be derived from different kinds of input data like the past activities, social links and preferences of the user and of other "similar" users (collaborative filtering) and/or the explicit description of the items and the preferences of the users (content-based filtering).

User similarity and recommendation in social networks : We consider the problem of discovering valuable content publishers in micro-blogging systems by providing efficient, topological and contextual user recommendations. Topology-based measures are good indicators for estimating user similarity and we propose personalized user recommendation scores which capture both the topological proximity and connectivity of publishers along with their topic authorities. In order to speed up the recommendation process we propose approximate algorithms based on landmarks [BD-C-4,BD-D-4]. Within this context, we also develop a novel block-based, workload-aware edge partitioning strategy for distributed user similarity computation [BD-C-2,BD-C-5,BD-D-2]. This work was done in collaboration with the Cédric-CNAM laboratory (PhD of R. Dahimene, Q. Grossetti and Y. Li) and was partially supported by the PEPS INS2I/INSMI 2015 FaSciDo "FAIT" (Finding Account of Interest on Twitter).

Online matrix factorization and tag recommendation : We consider matrix factorization algorithms in highly dynamic contexts with continuous rating streams where it is difficult to have an up-to-date recommendation model.Our solution reduces the lag-based recommendation quality loss by introducing biases tracking the user behavior deviation [BD-O-1]. We also propose an algorithm that takes into account the tags4 popularity and the user's neighborhood ratings. We propose a heuristic network traversal bound that enables on-the-fly recommendation computation with limited computation cost [BD-C-10]. We also propose a new dynamic method for improving the accuracy of top-k recommendation algorithms [BD-C-9]. Finally, we propose a new Geographical POI (Point of Interest) recommendation method which takes into account the delay and distance between POIs to better capture the set of POIs which might be worth to be visited together [BD-C-7] than other existing methods based on matrix factorization. This work is done in collaboration with the LTCI-Télécom ParisTech (PhD M. Gueye, PhD J.-B. Griesner).

Large-Scale text mining workflows : (B. Amann, H. Naacke, C. Constantin, B. Piwowarski) In the context of the ARESOS CNRS Mastodons project, the we have started to explore the usage of MapReduce based algorithms and platforms (Hadoop and Spark) for complex text and graph mining workflows. In collaboration with the Institut des Systèmes Complexes and IRISA (Université de Rennes), the goal is to design and implement scalable algorithms for building phylomemetic networks describing the evolution of science in form of temporal topic graphs. Scalability is achieved by combining data parallelism and incremental computation for topic extraction and topic alignment. After some preliminary results, this work is perpetuated in the ANR project EPIQUE which started in January 2017 under our coordination.

Representation learning and information access : (B. Piwowarski) The topic machine learning and information access is mainly represented by B. Piwowarski who enriched the team competencies for better understanding data science related problems and applications, and fosters the collaboration with the MLIA team around the theme of learning to project complex objects (e.g. text or graphs) into a continuous latent space for information access tasks : node classification [BD-MLIA-C-1], information retrieval, or word evolution. The quantum probability formalism, which is a way to represent objects in a continuous space, has been also exploited for summarization [BD-J-4].

Semantic IoT data integration : (B. Amann, M.A. Baazizi) In January 2014, we have started a new activity on Smart IoT applications as part of the two year exploratory project EBITA. The goal of EBITA was to

explore business and research opportunities between Fraunhofer (Germany) and UPMC concerning semantic data technologies developed in the DAPA department and the IOSB Fraunhofer Institute in Karlsruhe. During the two-year project duration, the team temporarily hosted the project leader (V. Tippmann) and two research engineers. EBITA also allowed us to finance an ongoing PhD on semantic IoT data enrichment and integration (F. Hannou). Taking the data generated by thousands of sensors at the Jussieu Campus as starting point, the goal is to develop new data preparation and integration methods for combining raw sensor data with contextual information (room occupation, energy consuming devices, outside temperature, ...) to enable higher-level semantic data exploration and analysis.

Schema inference for NoSQL applications : (M.A. Baazizi) Data formats like XML or JSON are more and more adopted because of their flexibility to encode semi-structured documents data. Compared to the relational model where data must strictly conform to a predefined prescriptive schema, JSON imposes no a priori schema simply because the structure of the data is unknown in advance or might change later. However, the absence of a schema also complicates the interaction with the data. First, formulating sound queries without a concise description of the underlying data is more difficult. The lack of a schema prevents query processors from performing a static analysis to check, for example, if a query uses nonexistent attributes. Finally, the lack of schema prevents from accelerating query processing by using well-known optimization techniques like wildcard expansion and data projection. We introduce an MapReduce based inference algorithm for extracting descriptive schemas from large JSON data-sets [BD-C-1]. The extracted schemas concisely and precisely capture the structure of the input data set. This work is done in collaboration with researchers from Paris-Dauphine University, University of Pisa and University della Basilicata, Italy.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journals	4 (0)	1 (1)	2 (0)	2 (0)	1 (1)	0 (0)	10 (2)
Books, chapters	0 (0)	0 (0)	1 (0)	2 (0)	0 (0)	0 (0)	0 (2)
Conferences	9 (0)	7 (3)	7 (2)	8 (1)	1 (4)	2 (0)	34 (12)
Software, demonstrations	1 (1)	0 (1)	3 (0)	0	0 (2)	0 (0)	4 (4)
Patents	0	0	0	0	0	0	0

$a(b)$ means a international and b national items. Journals and conferences are with proceedings. Software includes demonstrations and prototypes.

Sélection des produits et des activités de recherche

voir Annexe 4.

Faits marquants

The European Project (FP7) Scape has been very fruitful for our team : two PhD thesis and fruitful collaborations with European teams (e.g. national libraries of UK, Netherlands, Austria), two software components delivered to the project and used by other partners (academic and industrial). It also allowed to establish a tight scientific collaboration with the MLIA team (M. Cord) of LIP6.

The general objective of the two-year EBITA pilot project (jan 2015 - dec 2016) was to evaluate the opportunities for establishing a long term collaboration between the German Fraunhofer Institute and UPMC. For the BD team it was the opportunity to start a scientific collaboration with Fraunhofer IOSB in Karlsruhe on data processing issues generated by advanced Smart IoT applications. In particular, we started to collaborate on a shared prototype for smart grid management including complex data integration and analysis tasks. The project ended in 2016 but the team is continuing to work on this new topic.

During the evaluation period, the BD team established a new fruitful collaboration with the Institut des Systèmes Complexe (ISC-PIF, D. Chavalarias) and IRISA, Rennes (Z. Miklos) on large scale text mining applications. This collaboration started within the context of the CNRS MASTODONS action ARESOS and has been consolidated by a DIM PhD scholarship (S. Castillo) and the ANR project EPIQUE.

3. Analyse SWOT

Strengths : The main strength of the BD team is its complementary expertise in distributed databases, data modeling, data parallelism, query optimization and information retrieval. This expertise allows us to tackle a large spectrum of data-science related research problems which need a deep understanding of low-level database technologies (storage, indexing, distribution) and higher-level data analysis methods and algorithms. A direct positive consequence, linked to the current big data trend in science and industry, is the ease of hiring Bachelor and Master students and finding public and private fundings.

Weaknesses : Our main weakness is our difficulty to capitalize our research and development efforts and to make the results sustainable. This is due to the limited resources for doing the necessary engineering effort to package and maintain software, but also to the fast scientific, technological and industrial evolution of our research domains. This lack of sustainability combined with the difficulty to build competitive teams around more long-term projects makes it harder to compete with other top-level research groups and to produce the necessary human effort for participating in European projects and publishing more regularly in top level conferences (VLDB, SIGMOD).

Opportunities : There are opportunities for improving the scientific production through the participation in industrial and academic platforms. The EBITA project and collaboration with Fraunhofer was a first experience in this direction by hiring two full-time research engineers for facilitating technology transfer and software maturation. The new Tremplin Carnot INTERFACES also is an opportunity for making our research results more sustainable. Another important opportunity is our affiliation to the DAPA Data Science department, which is a clear asset for our visibility and for elaborating more ambitious data science research projects integrating the complementary expertise of the other DAPA teams.

Threats : Data-science and big data research evolves in a rapidly changing high-technology sector with important industrial actors (Google, Facebook) continuously hiring the best researchers in the domain. It becomes more and more difficult for academic research teams to compete with these actors. Another difficulty is to publish interdisciplinary and/or more application oriented research results in top-level conferences and journals. Finally, it gets more and more difficult to hire excellent PhD and post-doc students which are seduced by more attractive positions in industry.

4. Projet scientifique à cinq ans

Perspectives et objectifs

Data-driven science has become a new scientific paradigm (in completion to empirical, theoretical and computational science) because of the impact of information technology and the data deluge. The Database team will continue to leverage its long experience on large-scale data processing for finding new scalable access methods and algorithms for *information extraction, integration, exploration and analysis for the web and large-scale data science applications*.

Short-term perspectives

JSON schema inference : Structural and statistical metadata are essential for deciding optimal data storage policies or generating optimal query plans and helpful for assessing the quality of data-sets (data completeness, coherency). Our short-term goal is to adapt the current JSON Schema Inference approach to account for statistical information about the input data. The main challenges here are the definition of the extended schema model (semantics, quality criteria) and the implementation of efficient inference algorithms. The goal is also to address the precision-succinctness trade-off by devising parametric inference techniques allowing for inferring schemas of different precision.

IoT data curation and enrichment : We will continue to work on IoT data curation and integration. One main challenge is the semantic, spatial, temporal and structural heterogeneity of the various data sources to be integrated. Our goal is to define and develop a high-level declarative integration model and language which

allows the user to iteratively specialize a generic integration schema and generate mapping rules for combining raw sensor data with rich contextual information.

Scalable SPARQL query processing : Our goal is to extend our promising results on scalable SPARQL query processing with Spark by developing a full-fledged SPARQL query compiler and optimizer. We are currently looking for PhD fundings for working on this topic.

Long term perspectives :

Large-scale text mining workflows : The EPIQUE ANR project started in January 2017 and will represent a central part of our research activities during the next evaluation period. Our goal is to develop scalable topic extraction and alignment algorithms which are integrated in a global text mining framework. The main challenge is to adapt existing topic extraction and alignment methods by using data parallelism (MapReduce) to achieve horizontal scalability. In parallel, work will be conducted on end-to-end representation-based learning for text mining.

Social networks for malaria prediction : Late 2016 we started a new study on using social networks for malaria prediction in collaboration with UCAD. The idea is to use smartphone social networks to improve the malaria prevention. As users are geolocalized, the information about their status (infected or not) helps in predicting places where mosquito control must be proceeded with highest priority. Two master thesis are currently developed on the subject at both sides (LIP6 and UCAD).

Representations for information access : Two directions will be pursued - the first one continues existing works on Gaussian embeddings that are useful to represent the uncertainty of representation ; the second one is more prospective and aims at representing objects by graphs of embeddings (rather than single embeddings) which is necessary to represent complex objects like text.

User recommendation in social networks : The next step in our work on user recommendation in social networks is to gather deeper insight on graph-oriented computation in the recommendation context. Similarity computation at query evaluation is time consuming and too expensive to be stored in memory for large graphs. The recommendation model and computation must also consider the change of the underlying information (change in user opinions or of the social graph structure). We also intend to get a deeper understanding of the user behavior in social networks through homophily studies in order to incorporate user behavior into our recommendation algorithms Several optimizations will be considered like delta updating of the scores, grouping of messages, graph partitioning strategies and applying Monte Carlo strategies in a distributed setting.

Strategy for achieving these goals : Our strategy is continue our collaborations with other teams at the local (DAPA, LIP6) and the national level (Univ. Marne la Vallée, IRISA, Télécom, CNAM) and to enforce our contacts at the international level. We will also continue to invest in data science related research problems and push interdisciplinary collaborations with other scientific communities (human science, biology, health, etc.). This strategy follows the current trend of funding research projects which integrate data, techniques, concepts, and theories from two or more disciplines to solve problems whose solutions are beyond the scope of a single research domain. The team will evolve (one retirement, two potential promotions) and we intend to hire a permanent researcher during the next evaluation period.

EQUIPE LFI

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The research topics of the *Learning, Fuzzy and Intelligent systems* (LFI) team are related to the use of computational intelligence techniques, including machine learning and soft computing, to build intelligent systems. The team members conceive models to handle imperfect knowledge (e.g. imprecise, uncertain, subjective) both at a theoretical and a practical levels, to address application challenges. Real-world applications are conducted by the team, in a broad range of domains as multimedia, web-based image search engines, or temporal data.

Tableau des effectifs et moyens de l'équipe

In June 2017, the team is composed of 1 full professor (PR), 2 associate professors (MCF), 1 emeritus CNRS senior researcher (DR) and 14 PhD students. Among the latter, 4 are jointly supervised with a foreign university ("thèses en co-tutelle" with Kenya, Morocco and Poland) and 2 of them are also partly funded by means of an ANRT CIFRE with Thales ; 1 PhD is funded with an ANRT CIFRE with SevenHugs ; 2 PhD are funded with an ATER position (UPMC, Paris Diderot) ; the other 7 PhDs are funded with projects (LABEX SMART, FUI, or industrial projects with Arcelor Mittal, Total or AXA). Until the end of 2017, 5 PhD will defend their thesis.

During the period, the team has been weakened by the retirement of B. Bouchon-Meunier (DR CNRS), previous head of the team, and by several departures : M. Detyniecki (CR1, CNRS, HDR) went on a "détachement" position in the AXA group, M. Rifqi (MCF HDR) joined an econometrics team in the university Paris 2 where she is MCF, and N. Labroche (MCF HDR) obtained a mutation to the University of Tours.

In the following table, the row "Professors" stands for full professor or associate professor (MCF).

Workforces	2012	2013	2014	2015	2016	2017
Professors	5	5	4	3	3	3
Researchers	2	2	2	1	1	1
PhD students	8	7	9	10	12	14
Others		1				
Total	15	15	15	14	16	18

During the period, the team has participated to 11 funded projects (see Annexes 4 for the complete list).

Grants	2012	2013	2014	2015	2016	2017
International						
National	198 k€	135 k€	198 k€	198 k€	139 k€	105 k€
Industry	32 k€	32 k€	91 k€	109 k€	93 k€	133 k€
Others						
Total	230 k€	167 k€	289 k€	307 k€	232 k€	238 k€

(grants are broken down by year according to their duration).

Politique scientifique

The research of the LFI team belongs to several domains as machine learning, soft computing and approximate modelling to process, as input or output, uncertain and imprecise knowledge as well as subjective or sensory data.

In this setting, the team conducts research on Computational Intelligence approaches, based for instance on fuzzy logic, to model human reasoning and knowledge. Applications are used both to feed the theoretical research and to provide a way to evaluate the proposed approach. Applications with a high societal implication and at the heart of the numerical life are privileged. New thematics have been developed to address the challenge of temporal data processing, data stream, and subjective information in relation with affective computing.

The team is focused on both fundamental research and on industrial collaborations that provide real-world applications to evaluate models. The team produces mainly scientific papers in international conferences and journals. An important activity of scientific research is conducted with national and international academic partners and lead to the publication of joint papers (see publication list). The international visibility of the LFI team is satisfactory in the research community and can also be seen in the invitations both for co-supervising theses or to give plenary talks that LFI members receive. At the national level, the team is well introduced and very active in the soft computing community.

Industrial collaborations (direct through national projects) are numerous. They provide the main funding of the PhD theses supervised in the team and supply the necessary funds to attend conferences.

2. Produits de la recherche et activités de recherche

Bilan scientifique

The global underlying principle of the research developed by the LFI team is to combine various computational intelligence approaches, in particular machine learning and soft computing techniques, to propose models able to take into account human knowledge.

The research topics of the LFI team can be mainly structured around three main axes : supervised and unsupervised machine learning in presence of imperfect data, interpretability of AI and machine learning systems, and modelling of reasoning and knowledge. They are briefly presented below in turn, a more detailed example of a PhD thesis related to the subject is given for each of them.

Supervised and unsupervised machine learning in presence of imperfect information

Machine learning has traditionally been a very active research topic of the LFI team, the recent considered questions can be roughly summarised around the notions of temporality and contextualisation.

Temporality is a very general term that covers a wide span of challenges. The LFI team first addresses the issue of data that are available in the form of temporal series or data streams : it studies the proposition of efficient processing techniques, for their classification or their characterisation ; it also tackles the task of evolving data distributions and proposes soft computing approaches to deal with concept drift.

Contextualisation applies to the enrichment of machine learning systems, taking into account additional locally available information in the learning process or extracting local results, that do not apply globally but provide relevant complementary information, of interest to the user. In supervised learning, the LFI team contributions for instance include the proposition of original algorithms to build fuzzy decision trees exploiting the presence of class monotony, studied both at theoretical and experimental levels. In the unsupervised case, the LFI team addresses the task of subspace clustering, capable of simultaneously identifying (fuzzy) clusters and the subspaces they live in.

An example of the work of the team is illustrated by a thesis funded by the LABEX SMART, supervised jointly with the SYEL team of the LIP6 within the domain of affective computing. In this work, the aim is to build a user-independent, dynamic, embedded emotion recognition model exploiting physiological signals (EDA, EEG, ECG, EMG ...). The considered data and classes are typical examples of imperfect information, as they both present uncertainty and imprecision, as well as subjectivity. The system must be able to integrate new descriptors, as well as new classes (new mental states) and to detect abrupt changes or breaks without confusing them with noise. For this thesis, an experiment has been conducted, jointly with INSEAD, on a panel of 58 video-game players to build the physiological database.

Interpretability of AI and machine learning systems

The big data and data science challenges have recently attracted attention both from theoretical and applicative points of view. However, it appears that, if it is essential to build reliable and efficient systems, it is also crucial to offer interpretable systems and interpretable decisions. Indeed, it is often not sufficient for a system to predict correctly, it is also necessary to explain why and how a particular decision is taken. Building human-friendly machine learning models is an important issue in numerous application domains : increasing the interpretability of complex systems is essential and has become a major societal challenge.

In this setting, the LFI team first works on interpretable data science approaches, based on linguistic summaries, i.e. data description in natural language. The latter indeed provide concise and intelligible views of these data. The LFI contributions span over three main components : (i) the definition of original, rich and contextualised forms of linguistic summaries, for numerical data as well as for numerical series (bridging the gap to the previously mentioned research axis) ; (ii) the proposition of formal tools to guarantee the logical consistency of the built sets of sentences ; (iii) the development of efficient methods to extract such intelligible descriptions, including incremental algorithms and implementations.

Second, for the paradigm of supervised machine learning tasks such as classification, the LFI team works on the proposition of new frameworks to formally define human-friendly decisions : our aim is to gain a better understanding of the different facets of interpretability in a data science context. Taking into account the fact that interpretability of the inner structure of a complex system is of various kinds, we study it at different levels, including its global coherence and readability, the local validity of its components or the consistency of its outputs for instance. To do so, we propose to apply a cross-disciplinary approach, among other including cognitive science results.

The PhD thesis funded by Total belongs to this research topic : its aim is to extract interpretable knowledge from complex 3D structures to perform a classification task, the applicative domain is related to the study of the sub-soil within the framework of the discovery of hydrocarbons. In order to deal with the request for high interpretability of the obtained results, approaches based on the fuzzy prototype technique are exploited, as they offer readable outputs that can be validated by the domain experts. To improve the classification performance and efficiency, a combination with transfer learning approaches is currently being investigated.

Modelling of reasoning and knowledge

It is a well-established observation that classical logic does not model the way human beings reason and manipulate knowledge. The LFI team has a long-lasting expertise on non-classical logics, in particular weighted ones that have been shown to be more appropriate to model human-like reasoning. Its research topics in the last 5 years both deal with knowledge representation and manipulation.

Regarding knowledge representation, it proposes original approaches to model the imprecision naturally associated to linguistic terms, such as "*cheap flights*" or "*about 140*". The LFI contributions encompass the proposition of appropriate formal, as well as computational, models and their automatic instanciations in various contexts. The latter include data-driven approaches, for instance based on the the exploitation of observed flight prices, and user-driven approaches, based on cross-disciplinary collaborations, in particular with cognitive scientists. In particular, various original models to interpret approximate numerical expressions have been proposed and experimentally validated on empirical real data. The theoretical contributions on this topic in particular lead to practical contributions in the domain of flexible queries of numerical databases.

The domain of knowledge manipulation covers many types of issues, among which the LFI team for instance considers the combination of approximate numerical expressions in imprecise arithmetic calculations, or, at a formal logic level, various forms of logical inference, such as interpolative reasoning or graduality constrained implication.

Another major topic LFI deals with focuses of the manipulation of pieces of information, in a context of information scoring : it proposes a formal framework to model the trust building process, i.e. to measure the degree of trust that can be put in a piece of information. To that aim, the sequential integration of relevant dimensions, taking into account among others the information source, its content and the other available pieces of information, is performed in an extended variant of multivalued logic.

Another example of the work of the team is illustrated by a thesis funded by a "contrat doctoral" focusing on the case of graded belief representation and manipulation : a graded belief extends the notion of belief by

introducing a measure of the extent to which something is believed, requiring an extended formal framework as compared to modal logics classically used in this context. The aim of the thesis is to propose a weighted extension of modal logics, both from a semantic and axiomatic points of view. It then considers the specificity of the doxastic case and studies the relevance of several aggregation tools to model various manipulations of graded beliefs, in particular their conjunction, disjunction and negation.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	**Total**
Journaux/revues	6	3	1	3	5	2	20
Ouvrages / éditions	3	1	3		2		9
Chapitres de livres	1	2	1	4		3	11
Conferences	13	21	10	15	20	10	89
Logiciels							
Brevets							

Sélection des produits et des activités de recherche

See Annexe 4.

Faits marquants

Scientific contributions of the LFI team have been highlighted by the Computational Intelligence community. The team has been awarded two best student paper awards in conferences :

— at the international conference *FUZZ-IEEE'15 - IEEE International Conference on Fuzzy Systems*, Istanbul, Turkey, in August 2015 : "Oppositions in Fuzzy Linguistic Summaries", by G. Moyse, M.-J. Lesot and B. Bouchon-Meunier.

— at the national conference *LFA'16 - Conférence sur la Logique Floue et ses Applications*, La Rochelle, France, in October 2016 : "Optimisation proximale pour le subspace clustering flou" by A. Guillon, M.-J. Lesot and C. Marsala.

Notable facts are also the recognition of the scientific works conducted in the LFI team by the IEEE Computational Intelligence Society. B. Bouchon-Meunier received the 2012 IEEE Computational Intelligence Society Meritorious Service Award for her work dedicated to the promotion and the organisation of the IEEE Computational Intelligence Society. She was also awarded as a distinguished lecturer of the IEEE Computational Intelligence Society (from 2014-2016) that enabled her to provide a set of international seminars in various countries.

An important scientific results for the team is the first place obtained by S. Tollari among 6 international teams in the task "diversity" of the MediaEval Challenge in 2016, in Amsterdam. The task was dedicated to the problem of the diversity in results of web-based image search engines, in particular for multimedia, social and/or geolocalised information (texts, images, social networks, GPS,...).

The visibility of the team enables its members to participate actively in their community (see detailed list in Annexe 4) : to international program committees or as conference chairs. For instance, M.-J. Lesot served as program chair for IPMU'2016; B. Bouchon-Meunier as conference chair for WCCI'2012, and as technical chair for WCCI'2016. C. Marsala as COI chair for IFSA/SCIS'2017.

Moreover, the recognition of the scientific results and of specificity of the LFI team enables international collaborations that produced several scientific results (see the publication list). International collaboration are also highlighted by the four "co-tutelle" theses that have started with the University Hassan II in Casablanca (Morocco), the Strathmore University in Nairobi (Kenya), and two thesis in collaboration with the AGH institute of Krakow (Poland). These 2 theses are also partly funded by Thales.

3. Analyse SWOT

Strengths : The LFI team fully exploits the double thematic competence of its members both in the domain of machine learning and in the modelling of imperfect knowledge. In particular, the team is one of the main teams in France in the domain of fuzzy sets theory and is internationally well identified in its community. For

instance, the research history of the team in the domain of interpretability of the models places it in a very good position to develop new approaches in this hot topic in data science.

Weaknesses : The main weakness of the team is its size. During the last period, the team has been faced with several departures of senior researchers and no recruitment has been done to compensate for them. Therefore, it may become difficult for the permanent members of the team to face all the opportunities that could appear during the next five years.

Opportunities : In France, the LFI team can count on its multidisciplinary relationships, in particular with cognitive scientists from the LUTIN lab, and with industrial partners (AXA, Thales, Total,...). Several works have also been produced with other French academic teams (for instance LIRMM and IRISA that produce already some joint publications). Moreover, its participation in the LABEX SMART enables it to develop collaborations with local teams. The international visibility of the team enables it to count on a strong network of international collaborations (for instance, Germany, Spain, Italy, Poland, USA or Japan).

Threats : From a technical point of view, a threat lies in the high variability of the evaluation process of the proposed models. For instance, to evaluate the interpretability of proposed approaches and models, studies with experts or domain users should be conducted. In the works conducted jointly with the AXA team, domain users will be present to evaluate the results, but for other application domains, such panels will have to be found. Another threat is the evaluation of the whole results obtained by the team at the intersection of different domains. As LFI research is highly multidisciplinary (machine learning, cognition, logic,...) it may be difficult to position its works in specific conferences.

4. Projet scientifique à cinq ans

Scientific objectives

The scientific perspectives of the LFI team globally follow the same structure as its current activities and are presented using the same categories, although they go beyond them.

Machine learning and imprecise or subjective information. Research on temporal and stream data should be pursued in order to develop further our models that introduce imprecise modelling to smoothen data sets and introduce robustness for both the method to train the model and the consideration of forthcoming data.

On another axis, the team will also develop its work on diversity both in machine learning and in web-based image search engines. This work has been initiated in the team in the domain of ensemble learning and aim at proposing new approaches to reduce the size of an ensemble of classifiers. This topic is also important for our research for web-based image search engines that aims at proposing varied and diverse result rankings. This task offers a large set of real-world applications, in addition to application for search engines, the reduction of the size of an ensemble model in data science is an important approach to tackle the scalability of a model.

Interpretability of AI and machine learning systems. Research on interpretability will remain an important scientific research axis of the team. Nowadays, and for the forthcoming years, interpretability is one very active research topics and the LFI team aims at keeping its lead position in this topic.

Our goal still remains a study to obtain a better definition of interpretability and both to propose new frameworks to formally define human-friendly decisions and to apply them to the paradigm of machine learning models such as classifiers or clustering-based approach. In particular, our aim is to take into account the highly subjective component of interpretability which indeed depends on whom it is intended for (e.g. it depends on the recipient (e.g. the data expert in charge of understanding the system), his expectations and his knowledge which e.g. differ whether he is a mathematician or a computer scientist).

Modelling of reasoning and knowledge. Our scientific objectives include topics related to non-classical logics, so as to further develop our expertise in the domain. Among others, at a formal level, beyond the semantic and axiomatic study of the proposed weighted modal logics, the team will consider the question of automatising reasoning in this framework. In particular, the proposition of a tableau calculus and its implementation constitute a high priority perspective.

Our aim is also to take further advantage of this framework to study the relevance of its exploitation to model the relationships between beliefs and trust. Beside the logic point of view, this aim raises cognitive

questions to be answered through collaborations with cognitive scientists, as well as issues related to the choice of relevant weighted modalities to model attitudes such as sincerity or infallibility : the latter influence the trust put to an information source and to its outputs.

The study of conviction dynamics, related to the order in which pieces of information are gathered, also constitutes a challenging perspective of information scoring, that may be related to the domain of formal argumentation.

Strategic choices for the LFI team

The aim of the LFI team is to continue its collaborations at a local level (with other teams of the LIP6), at a national level (LIRMM, IRISA,...), and at an international level. The proposal of national and international projects will be pushed to develop these collaborations.

A key requirement for the team in order to carry out its scientific objectives is to recruit permanents researchers (UPMC or CNRS), which is the only way to maintain its research activities and to develop the new scientific thematics that have been highlighted. Moreover, it is a crucial necessity in order to enable the team to answer to the partnership demands it is faced with.

EQUIPE MLIA

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The central topic of MLIA is Statistical Machine Learning (ML) with an emphasis on algorithmic aspects and on applications involving semantic data analysis. The team is involved in several international and national academic collaborations and has developed a close cooperation with industrial R&D partners both through joint participation to projects and through bilateral contracts. MLIA is involved in international and national community service and at the local UPMC level is actively participating to both research and teaching administration.

Tableau des effectifs et moyens de l'équipe

Team composition

On June 2017, the team is composed of 9 permanent researchers : 3 professors, 5 maîtres de conférences, 1 chargé de recherche CNRS ; 27 PhD students (17 academic and 10 Cifres or industrial), 1 post doc and 4 research engineers. During this period, 4 researchers left the team : 2 promotions from maître de conférences to professor (M. Amini, N. Thome), 1 promotion on a CNRS Chair (N. Usunier), 1 professor mutation (T. Artieres). 2 maîtres de conferences were recruited in 2015 (O. Schwander and L. Soulier) and in 2017, 1 CNRS CR joined the team (Hichem Sahbi). This high number of promotions is an indication of the team strong dynamics. The number of PhDs has increased recently because of the strong industrial and academic demand for Data Scientists and the recent renewed interest for Artificial Intelligence and specifically for Deep Learning, one of the main thematic of the team. 23 PhD thesis have been defended over the 2012-2017 June period.

Workforces	2012	2013	2014	2015	2016	2017
Professors	10	7	7	7	9	8
Researchers	0	0	0	0	0	1
PhD students	22	24	17	18	19	27
Others	4	4	6	1	4	6
Total	36	35	28	26	32	42

TABLE 1.1 – MLIA members, 2012-2017

Projects and cooperation with Industrial partners The total budget of MLIA for the period is about 4.5 M Euros with roughly 2/3 coming from national projects (ANR, FUI, and special calls like CSOSG (security) or Big Data). MLIA has been engaged in 9 international projects (6 EU, 3 others). MLIA has a strong tradition of cooperation with industrial partners. Overall, we have been engaged in 12 collaborative projects (FUI or similar or bilateral). Industrial cooperation often comes through industrial thesis - a total of 4 collaborative thesis (mainly Cifre) have been defended during this period. Together with the LFI team we have launched in 2010 a joint lab (CLEAR) with a division of Thales. This lab. has been re-conducted three times for a period of 3 years. The Complex System team of LIP6 has joined the lab. in 2014. More details are provided in the annex.

Politique scientifique

The central focus of the team is Statistical Machine Learning and Deep learning. We develop algorithmic and theoretical contributions together with applications in selected domains. Concerning the algorithmic and

Grants	2012	2013	2014	2015	2016	2017	Total
International	144.2	158.7	94.3	27.0	241.3	129.5	795.1
National	1035.7	556.6	504	577.7	480.9	226.5	3381.3
Industry	89	119.8	67	61.4	148.5	58.5	544.2
Others	0	0	0	0	0	0	0
Total	1268.9	835	665.3	666.3	870.7	414.4	4720.6

TABLE 1.2 – Grants, 2012-2017 (k€)

theoretical contributions, the main domains explored during 2012-2017 have been 1) the development of Representation learning and Deep Learning algorithms for a variety of ML problems ; this has been an emerging field during this period, 2) the prediction of complex structured outputs which covers several fundamental ML problems for predicting structured data, 3) Sequential and Reinforcement learning which after a long period of pure academic and theoretical development has emerged as a field with a strong potential in several application domains. Concerning the applications, the team focuses its efforts on three main topics : Computer Vision ; Natural Language processing, Information Retrieval and recommendation ; Complex Data Analysis which covers the analysis of a variety of data collected from complex environments. These different research directions are detailed below.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Representation learning and Deep Learning (*All*) : Representation Learning emerged over the last ten years at the confluence of different domains and is best known today through the popular domain of Deep Neural Networks. MLIA has always have a strong position on this domain. A large part of the team research is today centered on this field and MLIA is probably one of the most prominent teams on Deep Learning in France. Over this period, we have investigated Representation Learning for structured and dynamic data, with applications in Computer Vision, Natural Language Processing, Social Data Analysis and Recommendation. This topic is then transverse to many MLIA activities and most PhDs currently engaged in the team are working on a topic related to Representation or Deep Learning.

Structured outputs (*T. Artieres, N. Baskiotis, P. Gallinari, N. Thome*) : learning to produce complex outputs encompasses many fundamental and application ML problems. A first research direction explored during this period has been "'Extreme Classification"', a new challenge consisting in classifying millions of classes. MLIA has co-organized 4 international challenges (LSHTC - Large Scale Hierarchical Text classification http ://lshtc.iit.demokritos.gr/), including a Kaggle competition) and workshops at international conferences (ECML, WSDOM, ICML) on this topic from 2011 to 2014. MLIA has also participated to a EU project (BIOASQ, http ://bioasq.org/) and was in charge of the development of a platform for large-scale biomedical semantic indexing of textual documents ; the platform is still running. 2 thesis have been defended on this topic. Another line of research is classification on graphs with an emphasis on heterogeneous graphs where nodes and edges may be from different types. Although graph node classification has been popular in ML, this was an entirely new topic for which we proposed new classification methods. This was motivated by applications in the domain of social networks. We have also investigated classification in graphs for social networks and image segmentation. In a Franco-Canadian international ANR project "Deep Vision", we are combining deep learning and structured prediction. 3 PhDs (M. Cisse, Y. Jacob, R. Puget) defended on this topic.

Sequential and Reinforcement learning (*L. Denoyer, S. Lamprier*) Many machine learning problems can be revisited as sequential learning problems. This allows one to define new learning methods, and to extend existing models to more complex tasks. We have investigated the development of new sequential models for structured output prediction, graph processing, budgeted learning, adaptive information acquisition, conditional computation and attention methods. We have organized one workshop on this topic at ICML2013 [1]. 1 thesis (G. Dulac-Arnold) has been defended during the period, 1 in July 2017 (G. Contardo), others are currently ongoing.

1. https ://psm.lal.in2p3.fr/

Theoretical and algorithmic contributions are developed together with practical applications in selected domains at MLIA. We review below the contributions for the main application domains of the team.

Computer Vision (*M. Cord, N. Thome*) : MLIA has a strong research activity on learning for Computer Vision. We have developed theoretical and algorithmic frameworks for visual pattern detection and recognition. To get useful models, we have explored computer vision-based approaches (CVIU 2013, ICCV 2015, CVPR 2016), bio-inspired modeling (IEEE IP 2013, CVPR 2013), and deep learning strategies (NIPS 2013, TNNLS 2014, CVPR 2016, 2017) for many years. We have several contributions to identify how these approaches share some general hierarchical data representations. We have exhibited some interesting sparse representations, investigated dictionary learning, and proposed to learn deep nets with a new combination of unsupervised and supervised strategies. We also focused on specific aggregation functions (pooling) for theses architectures because we consider that it is a key to efficiently encode invariance in nets. All these developments give to the team a strong position in the current area of deep convolution neural nets for visual information processing. We also developed original approaches for metric learning and semantical indexing for visual information retrieval tasks. We participate to several national and international projects and collaborations (ANR, European IP, Brazil, Canada). 6 PhDs were defended on this topic during the period (S. Avila, M. Chevalier, H. Goh, M. Law, C. Le Barz, D. Pitzalis). H. Goh was one of the first thesis in France on Deep Learning for vision.

Natural Language Processing and Information Retrieval, Recommendation (*M. Amini, P. Gallinari, V. Guigue, L. Soulier, N. Usunier*) : MLIA has been involved in the text community, mainly from an information retrieval (IR) perspective for years with contribution on learning to rank and on information retrieval for semi-structured data. The recent recruitment of L. Soulier has reinforced the IR orientation, impulsing a new line of research on Deep Learning for IR (ICTIR 2017, AIME 2017) and natural language understanding (CLEF 2017). This is done in a close collaboration with B. Piwowarski of the BD team. MLIA has also developed a research for about 10 years on recommender systems with both theoretical contributions on ranking criteria and practical contributions on the development of multimodal systems for joint review, text polarity and recommendation prediction. This line of research has been quite successful with 3 best paper prizes, 1 for theoretical contributions on ranking losses (NIPS 2012), 2 for algorithmic and experimental contributions (Recsys 2012 which is the premier conference on recommender systems and CORIA 2014 the French conference on IR). 6 PhDs defended during this period on the topic (D. Buffoni, C. Calauzenes, A. Fakeri, M. Poussevin, B. Pradel, A. Rafrafi).

Complex dynamic data analysis (*N. Baskiotis, L. Denoyer, P. Gallinari, V. Guigue, S. Lamprier, O.Schwander*) We have developed new models for the analysis and the exploitation of complex data streams, motivated by applications in different domains. A first topic has been the modeling of diffusion processes applied to information propagation in social networks. We have developed an entirely new perspective on the problem by modeling the diffusion process onto a latent continuous space instead of considering as usual a discrete diffusion process on a graph. This change of paradigm allows one to solve this type of problem using continuous optimization methods instead of discrete simulation. Another line of research also inspired by the analysis of social streams is the development of algorithms for collecting information from social sources using a limited resource budget. Several variants of this problem have been developed based on bandit formulations. A recent direction of research is the development of spatio-temporal prediction methods based on dynamic graphical models. The goal is to propose generic families of predictive methods for this problem which is still largely unexplored in the machine learning community. On a more applicative perspective, we have developed a research direction for analyzing the mobility problems in the transportation domain. This work has been carried out in cooperation with different actors like Thales, Ile de France Transportations (STIF), and Vedecom. All these research directions concern the analysis of dynamic data with temporal and sometimes spatial dimensions. MLIA has been coordinating for 4 years (2012–2015) a pluridisciplinary project "ARESOS" in the context of the CNRS MASTODONS call dedicated to big Data Analysis. This project focused on the analysis of social data was the only SHS-INS2I project accepted in this call. It gathered partners from 10 labs in disciplines like Computer Science, Cognitive Science, Natural Language Processing. Overall, for this topic of "complex dynamic data analysis", we have been able to develop original approaches on different problems by formulating them from a representation learning perspective. We believe that we have been pioneer in several of these directions. We also analyze data with a very strong geometric structure such as symmetric positive definite matrices coming from time series (project PRESED) or radar observations. These problem are addressed with methods both from information geometry and deep learning. 3 PhDs defended on this topic (L. Aurelien-Gauthier, S. Bourigault,

T. Gisselbrecht)

National/ International Collaborations MLIA has developed close international cooperation with China, Brazil, Italy, Singapore involving co-publications and researcher exchanges. It participates to joint projects with different countries (Belgium, Canada, Germany, Greece, Italy, Saudi Arabia, Spain, Switzerland). In France MLIA has been collaborating with teams in Grenoble, Lille, Marseilles, St Etienne and Toulouse. Over the period, MLIA has been involved in 9 international and 27 national projects.

Teaching MLIA is strongly involved in the Data Science Master cursus launched in 2014 (DAC speciality of the CS master), with 9 courses taught by team members, and L. Denoyer co-responsible for this cursus. The master cursus is closely related to the research topics of MLIA and of the DAPA department and we host each year several internships at the master level. MLIA also participates to the Big Data certificate of the mathematical master cursus (2 courses) and to the master of Statistics (2 courses).

National and local implication in scientific animation

— Member ANR committee CES38 2015, Contint 2011, 2013 (M. Cord)
— Member PEPS jury CNRS 2011 et 2012 (M. Cord)
— co-chair of action IRIM (Indexation et Recherche d'Information Multimédia) from GdR ISIS, 2012-2017 (M. Cord)
— Chargé de mission at CNRS INS2I, 2015– (M. Cord)
— Member of the France Artificial Intelligence thinking group reporting to the French Research ministry (M. Cord, L. Denoyer)
— Member ANR Comité d'évaluation scientifique CE23 (Data and HPC) 2016 (L. Denoyer)
— In charge of Open Research/Open Data/Open Education prospective for UPMC, reporting to the president of the university, 2016-2017 (L. Denoyer)
— Director LIP6, 2008-2012 (P. Gallinari)
— President Scientific commitee faculty of engineering, 2013 – (P. Gallinari)

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journals/revues	7	7	6	6	6	3	35
Books	0	0	1	0	0	0	1
Conferences	19	17	18	18	32	7	111
Softwares	0	0	0	1	2	1	4
Patent							0

TABLE 1.3 – MLIA publications in 2012-2017

Numbers for 2017 are acceptations at the date of May 2017.

Sélection des produits et des activités de recherche

Distinctions

— AFIA prize 2016 for a thesis in AI (national prize) for Marc Law.
— Best thesis prize (SIBGRAPI 2013) for Rodrigo Minetto cotutelle with university Campinas, Brazil

Best paper prizes

— Best paper CORIA 2017 (national), Bourigault S., Lamprier S, Gallinari P., Apprentissage de representation pour la detection de source de diffusion dans les reseaux, CORIA 2017.
— Best student paper ICONIP 2016 (international), Contardo G., Denoyer L., Artières T., Recurrent Neural Networks for Adaptive Feature Acquisition.
— Best paper CORIA 2014 (national), Poussevin M., Guàrdia-Sebaoun E., Guigue V., Gallinari P. : Recommandation par combinaison de filtrage collaboratif et d'analyse de sentiments. CORIA-CIFED 2014 : 27-42.

— Honorable mention : RECSYS 2012 (international), Pradel B., Usunier N., Gallinari P. : Ranking with non-random missing ratings : influence of popularity and positivity on evaluation metrics. RecSys 2012 : 147-154.
— Honorable mention : NIPS 2012 (international), Calauzènes C., Usunier N., and Gallinari P., On the (Non-)existence of Convex, Calibrated Surrogate Losses for Ranking, NIPS 2012.

Project Coordination MLIA has coordinated the following projects

— ARESOS - Reconstruction, Analyse et Acces aux Données dans les Grands Reseaux Socio-Semantiques (CNRS Mastodons call) project for 4 years. 10 research teams from different disciplines involved.
— MUSTER EU Chist Era project - partners : KU Leuven, ETH Zurich, University of the Basque country.
— 2 ANR projects : LOCUST, Class-Y.

Besides, MLIA actively participates to a joint team (CLEAR) with a Thales Lab. During this period, the joint activities mainly concerned research on diagnosis, algebraic models of recommendation and transport.

Grants

— 2016 MLIA was selected together with 22 European teams, as a recipient of the Facebook AI Research (FAIR) partnership program for providing academic labs GPU servers for promoting AI and Deep Learning. MLIA is one of the 3 French teams selected by this program. Note that 3 former PhD of MLIA are currently researchers at FAIR. Considering the extremely highly selection level of FAIR, this is also an indicator of the quality of the PhDs from the team.
— 2016, Technicolor grant (via UPMC Foundation), 50k euros for supporting the research activity on Deep Learning for vision.

Scientific achievements

We describe below a few representative papers from our scientific publications.

[MLIA-J-7] In this paper, we study visual modeling architectures that combine the Bag of Words framework with the representational power and adaptability of learning deep architectures. Local gradient-based descriptors are encoded via a hierarchical coding scheme composed of spatial aggregating restricted Boltzmann machines (RBM). For each coding layer, we regularize the RBM by encouraging representations to fit both sparse and selective distributions. Supervised fine-tuning is used to enhance the quality of the visual representation for the categorization task. The low-level representations of descriptors that were learned using this method result in generic features that we empirically found to be transferrable between different image datasets. The representations learned are compact and the model's inference is fast compared to sparse coding methods. Further analysis reveals the significance of supervised fine-tuning when the architecture has two layer of representations as opposed to a single layer.

[MLIA-J-12] This paper introduces the use of sequential decision algorithms for machine learning problems usually solved via direct loss optimization algorithms (e.g. gradient methods). We have explored this idea through several publications. The focus of this specific paper is on learning sparse representations and cost sensitive classification. In supervised classification, data representation is usually considered at the dataset level : one looks for the "best" representation of data assuming it to be the same for all the data. We proposed in this paper a different approach where the representations used for classification are tailored to each datum in the data space. This representation is obtained by way of a sequential decision process that sequentially chooses which features to acquire before classifying a particular point ; this process is learned through algorithms based on Reinforcement Learning. The method extends easily to a whole family of sparsity-related problem which would otherwise require developing specific solutions. This is the case in particular for cost-sensitive and limited-budget classification, where feature acquisition is costly and is often performed sequentially.

[MLIA-C-7] We study in this publication the structure of very deep architectures for image classification, both for supervised learning and for a weakly supervised framework. In particular, we highlight the role of pooling as local and/or global aggregation operator. This pooling information is definitively important to combine geometric information with the recognition process in Computer Vision. We show the interest of what we call fully convolutional nets (FCN), where the pooling step is achieved at the last layer of the deep network. With these FCN architectures and with dedicated pooling functions, we study how detection, localization and classification may be analyzed in an unified learning framework. We obtained State-Of-the-Art performances for two years on several computer vision benchmarks (CVPR 2016 and our recent extension in CVPR 2017).

[MLIA-C-17] Analyzing and modeling the temporal diffusion of information on social media has mainly been treated as a diffusion process on known graphs or proximity structures. The underlying phenomenon re-

sults however from the interactions of several actors and media and is more complex than what these models can account for. We introduced in this paper a fundamentally different approach to this problem by 1) developing a purely data intensive approach, 2) mapping the diffusion problem in a continuous latent space instead of considering discrete models. The proposed approach develops an entirely new vision for this problem. It possesses several unique characteristics. Since its parameters are directly learned from diffusion cascade samples without requiring any additional information, it does not rely on any pre-existing diffusion structure. Because the solution to the diffusion equation can be expressed in a closed form in the projection space, the inference time for predicting the diffusion of a new piece of information is considerably reduced compared to discrete models.

[MLIA-C-18] This is one of our papers attacking the problem of Extreme Classification (classification with a very large number of classes), which has been one of the topics of the team with a strong implication in international collaboration and in the organization of international challenges. The paper presents an approach based on a reduction to binary classification in which label sets are represented by low dimensional binary vectors. This representation follows the principle of Bloom filters, a space-efficient data structure originally designed for approximate membership testing. The approach is provably robust, has sublinear training and inference complexity with respect to the number of labels, and compares favorably to state-of-the-art algorithms on two large scale multilabel datasets.

3. Analyse SWOT

- *Strengths ;* MLIA benefits from a diversity of skills, ranging from pure statistical machine learning and deep learning which is the core competency of the team, to application domains like text and vision. This is a relatively unique situation in France which allows us to position MLIA both on applicative and on ML communities. The team has an historical research background and international recognition on Neural Networks and benefits from the current enthusiasm of young researchers and others for Deep Learning.

- *Weaknesses ;* Because of the many opportunities in this domain, it is mandatory to focus on well chosen research directions both for research and for RD cooperations. Over the last few years, 4 young senior researchers have been promoted and left the team, which considerably reduced our potential and the number of HdR. Team researchers will be encouraged to defend their HDR. There is also a strong demand for teaching activities on this topic, the team members are heavily solicited and the potential of the team on this aspect has reached its limits.

- *Opportunities ;* The field is today extremely popular. Together with the launch of a successful Data Science Master Degree at UPMC, this has attracted bright students to the team these last years. Many of them also come from other cursus than UPMC masters. There is currently a very strong industrial demand in this domain and our PhDs really benefit from this context.

- *Threats.*

Brain drain towards industry is a major concern today in our domain. It is then extremely difficult to recruit bright PhDs on university positions, and many ML teams all over the world are decimated by this draining. Also the field has become extremely competitive with the recruitment of large groups of researchers by some major actors (Google, Facebook, Amazon, etc) and by top universities all over the world. This changes the nature of research in our field. Research today requires large investment in human resources and in equipment, which was not the case up to some years ago. We rely on our capacity to attract such investments and also on the potential perspicacity of the university and CNRS to support this research direction.

4. Projet scientifique à cinq ans

Over the last period, Representation Learning and Deep Learning have become central themes in the team. We envision an increased investment on these thematics together with the development of research directions which provide a clear differentiation w.r.t. large research groups at IT companies. We have planed the development of new topics in relation with the research already developed over the last period :

Multi-modal perception Man machine interactions require perception abilities in domains like vision, natural language processing, emotion and sentiment analysis, trace analysis, multi-sensors analysis, etc. We will develop research directions for the analysis of these different types of signals with an emphasis on multimodality. If we consider for example vision or language, each modality only offers partial information. Key

challenges for exploiting both modalities require the ability to ground abstract language in visual perception and symmetrically to abstract visual information in language. Our goal is to attack such key challenges for different tasks requiring the integration of multiple interaction modalities .

Unsupervised learning and learning with weak supervision The main successes of ML up to now concern learning with supervision (classification, ranking or regression). There are two problems with this approach, first data are often cheap to collect but supervision is often expensive or impossible, and second future AIs should be able to learn by themselves. We will develop several directions relevant to this topic : Deep Reinforcement Learning, Transfer and Distant Learning with applications in perception tasks (e.g. Natural Language Processing and Vision). In the field of Representation Learning, we are planning to investigate new weakly supervised representation learning methods. One key problem in this domain is the problem of information disentanglement i.e extracting independant latent factors in the representation learning process. One first article has been proposed this year in collaboration with Facebook Research (FAIR) on a computer vision process. We are currently investigating how this method can be extended to various applications like multi-view learning, stream learning, or time series.

Statistico-mecanical models and spatio-temporal processing There are two main paradigms for incorporating data information into predictive models. One is inherited from physics and rely on a mathematical modeling of explicit knowledge about the phenomenon to be analyzed (e.g. climate models). Data are then used to estimate some model parameters. The second one is the data intensive paradigm were general predictive models are trained from data. We plan to analyze the frontier between these two paradigms and their potential reciprocal enrichment for the modeling of spatio-temporal signals. One application field could be climate in cooperation with LOCEAN (Laboratoire d'Océanographie et du Climat) at UPMC.

Human-Machine interaction Our objective here is to investigate the problem of learning models able to interact with humans. This interaction can take various forms like natural language or demonstrations for example. We plan to develop new Deep Reinforcement Learning models specifically for learning how a system can interact with a user, with application to active learning, intelligent services in autonomous vehicles (collaboration with Renault), or dialog systems.

These developments represent new orientations for the research topics described before. In the next period we then plan to continue research on Vision, NLP and Information Retrieval, Complex Dynamic Data Analysis.

Statistical machine learning and deep learning are among the most competitive research topics in computer science today. Key requirements for carrying out a high quality academic research and more generally for developing Data Science at LIP6 are 1) the recruitment of permanent researchers to answer the high demand of academic and industrial partners on this domain, 2) access to computing facilities like large GPU clusters. One of the very first recommendation of the Scientific Advisory Board (SAB) that examined LIP6 in 2016 was "While it is important to maintain and improve the traditional CS areas, it is even more important to be allow sufficient increase in areas of CS that are currently of high demand and interest, e.g., machine learning." This vision requires developing a true scientific policy at LIP6 and UPMC.

DEPARTEMENT **DESIR**

1. Présentation du département

The department's research activities fall within artificial intelligence, decision theory and operational research, as suggested by the name DESIR standing for (Decision, Intelligent Systems and operations Research). They concern various related topics such as modeling and optimization of systems, algorithms and heuristics for problem solving, combinatorial optimization, computational decision support, adaptive agents and multi-agent systems, interactive systems and training, learning. We aim to cover a broad spectrum of activities including formal foundations (mathematical modeling, axiomatic analysis, optimization methods), computational issues in problem solving (complexity of problems and algorithms, design and implementation of exact methods, approximation algorithms with performance guarantees), the development of intelligent systems (adaptive agents, multi-agent systems, decision support systems, learning systems) and actual applications in industrial and educational contexts. The department currently includes 41 permanent researchers and professors (including 6 CNRS full-time researchers), and 28 PhD students. We are organized in four teams covering a continuum of topics ranging from Operational Research to ArtificiaI Intelligence, namely RO (Operational Research), Decision (computational decision-support), SMA (Multiagent Systems) and MOCAH (Models and Tools in Knowledge Engineering for Human Learning). More precisely, the main keywords of the teams are the following :

OR team : Operational Research, algorithms and heuristics, combinatorial optimization, scheduling, network routing

Decision team : computational decision-support, multicriteria optimization, reasoning and decision-making under uncertainty and risk, computational social choice, context, optimization of large scale systems.

SMA team : muliagent systems, coordination, distributed decision-making and multiagent planning, dialogue and interaction, simulation.

MOCAH team : interactive environments for human learning, learner modeling, metadata, cognitive modeling.

2. Politique scientifique

The DESIR department has a strategic position at the interface of Artificial Intelligence and Operational Research and, beyond the developement of our favorite research activities in both domains, we aim to foster positive synergies between the two research areas. The presence of people having different backgrounds (AI, OR, Decision Theory) in the department is indeed an opportunity to develop fruitful interactions exploiting the complementarity of the teams. For example we develop research and projects at the intersection of combinatorial optimization and decision theory (algorithmic decision theory) or at the intersection of multiagent systems, social choice and game theory (computational social choice, algorithmic game theory), and also at the intersection of knowledge representation and serious games (see Figure 1.1 for an overall picture of the interactions). The members of DESIR share some knowledge and information about their respective research topics and activities through a periodic *department seminar* organized at LIP6. This seminar is mainly dedicated to scientific presentations made by local permanent researchers. The department meetings are also the opportunity to introduce, once a year, the new PhD students, with a short presentation of their main research topic and their thesis objectives. In addition to this seminar, meetings of the four team leaders are organized to share information, coordinate actions and investments, or discuss opportunities and priorities for recruitement. We can also mention the existence of a reading group that sometimes proposes a collective study and discussion of a scientific book of general interest. The benefit of these interactions is illustrated by joint publications in

various domains such as multiagent based serious games [1], tutoring systems and user modeling [2,3], algorithmic game theory [4,7,8], decision theoretic planning [5], optimal argumentation [6], viability analysis [9], scheduling [10] and by several co-supervisions of PhD students.

The teams are also involved in joint academic or industrial projects. Let us mention for instance the ANR projects GUEPARD (Exact and Approximate Algorithms for Multiobjective Combinatorial Optimization), Coccorico (Computation, Communication, Rationality and Incentives in Collective and Cooperative Decision Making) and COCA (combinatorial Optimization with Competing Agents), involving colleagues from two or three different teams of the department each time. Other cooperations occur in private contracts for industrial applications, for example in the Dem@Factory project concerning the control and optimization of massive digitization processes. We also cooperate in LIP6 projects such as the OPT project on scheduling plans of tasks for multi-sensor multi-functions systems.

Finally, we also cooperate in Education by jointly teaching AI, decision theory and OR in the ANDROID computer science master program at UPMC. In this program developed in cooperation with some colleagues of the ISIR lab, master students are trained in multiagent systems, robotics, optimization and decision making, and interactive tutoring systems.

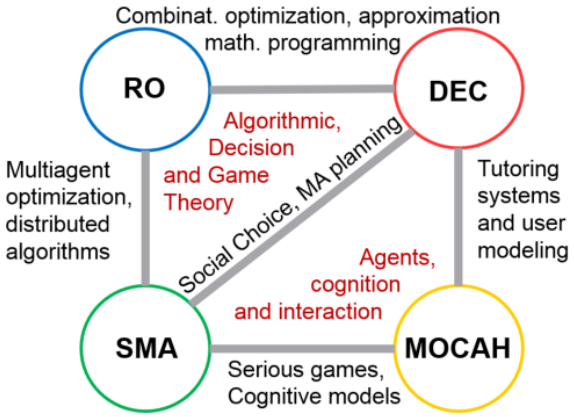


FIGURE 1.1 – Main interactions in DESIR

References

[1] Jean-Pierre Briot, Marta de Azevedo Irving, José Eurico Vasconcelos Filho, Gustavo Mendes de Melo, Isabelle Alvarez, Alessandro Sordoni, Carlos José Pereira de Lucena ,“Participatory Management of Protected Areas for Biodiversity Conservation and Social Inclusion : Experience of the SimParc Multi-Agent-Based Serious Game", In Multi-Agent-Based Simulations Applied to Biological and Environmental Systems, IGI Global, 295-332, 2016.

[2] Yang Chen, Pierre-Henri Wuillemin, Jean-Marc Labat : Bayesian Student Modeling Improved by Diagnostic Items. Intelligent Tutoring Systems 2014 : 144-149.

[3] Yang Chen, Pierre-Henri Wuillemin, Jean-Marc Labat : Discovering Prerequisite Structure of Skills through Probabilistic Association Rules Mining. EDM 2015 : 117-124.

[4] Bruno Escoffier, Jérôme Monnot, Fanny Pascual, Olivier Spanjaard : Truthful Many-to-Many Assignment with Private Weights. CIAC 2013 : 209-220.

[5] Emmanuel Hadoux, Aurélie Beynier, Paul Weng : Solving Hidden-Semi-Markov-Mode Markov Decision Problems. SUM 2014 : 176-189.

[6] Emmanuel Hadoux, Aurélie Beynier, Nicolas Maudet, Paul Weng, Anthony Hunter : Optimization of Probabilistic Argumentation with Markov Decision Models. IJCAI 2015 : 2004-2010.

[7] Anisse Ismaili, Evripidis Bampis, Nicolas Maudet, Patrice Perny : A study on the stability and efficiency of graphical games with unbounded treewidth. AAMAS 2013 : 263-270.

[8] Anisse Ismaili, Evripidis Bampis, Nicolas Maudet, Patrice Perny : How the number of strategies impacts the likelihood of equilibria in random graphical games. AAMAS 2014 : 285-292.

[9] Sophie Martin, Isabelle Alvarez, Jean-Daniel Kant : Micro/macro viability analysis of individual-based models : Investigation into the viability of a stylized agricultural cooperative. Complexity 21(2) : 276-296 (2015).

[10] Yasmina Seddik, Christophe Gonzales, Safia Kedad-Sidhoum : Performance guarantees for a scheduling problem with common stepwise job payoffs. Theor. Comput. Sci. 562 : 377-394 (2015).

EQUIPE DECISION

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The *Decision* team gathers computer scientists working at the crossroad of decision theory, artificial intelligence and operational research. The team, initially focused on mathematical preference modeling and decision support systems, has progressively extended its range of activities to develop the computational aspects of decision support, motivated by the increasing needs of computational models for decision making on combinatorial domains. Patrice Perny, professor in the team since 2002 has developed a small group on algorithmic decision theory, Professor Michel Minoux joined the team in 2004 with a small group working on combinatorial optimization and operations research, and Christophe Gonzales, professor in the team since 2008, developed with a small group the activity on the algorithmic of reasoning and decision making under uncertainty. These groups intersect each other and cooperate to address the new challenges posed by computational decision support. In parallel, the initial activity on intelligent decision systems has been maintained by Professor Patrick Brézillon.

Tableau des effectifs et moyens de l'équipe

In 2012, we welcomed Paolo Viappiani (CR1 CNRS) with the aim to develop our activity on preference elicitation and preference learning. In the same year, we also recruited Thibaut Lust as "Maître de conférences" to reinforce our activity on multiobjective combinatorial optimization. Unfortunately we also had several significant departures. In 2014, Paul Weng left for a position of Assistant Professor at SYSU-CMU (China), and in 2016, Thibaut Lust took a senior researcher position at Dublin Institute of Technology. Isabelle Alvarez who was part-time researcher in our team, delegated by IRSTEA, recently moved to the Institute of Complex Systems. Moreover, Professor Michel Minoux became "emeritus" in 2012 although he actively continues to lead the group on systems modeling and optimization, and Professor Patrick Brézillon will retire next year.

Workforces	2012	2013	2014	2015	2016	2017
Professors	8	9	9	8	7	7
Researchers	1	2	2	2	2	1
PhD students	11	10	8	10	10	9
Others	1	1	1	1	1	1
Total	21	22	20	21	20	18

The total grant of the team for the period 2012-2017 was about 1.5 Meuros. It mainly comes from European and national research contracts as well as private research and development contracts with industry. The team is or was involved in several academic national projects (supported by the ANR) such as GUEPARD (leader), LARDONS, COCA, CoCoRICo-CoDec (see Annexe 4). Moreover, the team also has a strong commitment toward industrial transfer, project and contracting. As such, we have been part of two European projects (FP7 MIDAS and H2020 SCISSOR), the ANR project DESCRIBE, a Digiteo project (CORBOS), and we have various partnerships (e.g., with Airbus, EDF), several CIFRE contracts (e.g., with IBM, Akheros, France Telecom) and industrial contracts (e.g. with IRSN). For more details see Annexe 4.

Grants (Keuros)	2012	2013	2014	2015	2016	2017
International	33.88	71.81	68.52	200.49	145.00	139.25
National	142.63	165.61	121.81	118.95	52.87	34.98
Industry	19.26	17.25	28.39	37.53	22.48	16.06
Others	8.99	0	0	0	0	0
Total	204.76	254.67	218.73	356.97	220.35	190.30

Politique scientifique

Our research activity aims to develop formal models and algorithms to help humans or machines to make better decisions in complex environments. We work on probabilistic reasoning, decision-making and optimization under uncertainty and risk, multiobjective optimization, collective decision-making, and on the optimization of large systems (telecommunication, transport, energy). A significant part of our activity is dedicated to theoretical studies concerning the mathematical properties of decision models, the formal representation of decision and optimization problems, the analysis of the complexity of problems and algorithms, the design of exact algorithms or approximate algorithms with performance guarantees. Our researches fall within Artificial Intelligence and Operations Research and we publish in both communities. Our strategy targets top international AI conferences and selective OR and AI international journals. This theoretical part of the work is motivated by the aim to provide practically efficient tools to support decision making in actual applications. Hence most of our studies concern decision or optimization problems that are abstractions of real complex decision situations that human or artificial agents may have to face in practice. We develop computational models for decision support (rational preparation of important decisions, recommender systems) and for automated reasoning and decision making (autonomous decision agents). More precisely, we can distinguish three complementary and interacting lines of research in the team :

Algorithmic Decision Theory (ADT). We work on computational models for decision-making and planning under uncertainty and risk, preference aggregation and multiobjective optimization, social-choice and collective decision-making, preference elicitation and learning, context-based decision-making and explanation.

Graphical Models (GRM). We work on inference and learning in Bayesian networks. We develop fully object-oriented Probabilistic Relational Models (PRM) and work on both modeling and structured inference algorithms. We also study the use of graphical models for the compact representation of preferences.

Modeling and Optimization of Systems (MOS). We work on new types of relaxations for the exact or approximate solution of hard integer optimization problems arising in many industrial applications such as telecommunication networks, logistic and transportation networks, computer-aided design of VLSI components. We also investigate robust optimization models in multistage decision making under uncertainty.

The team is one of the partners and founding members of the international research group on Algorithmic Decision Theory (International GDR ALGODEC created in 2011 and supported by the CNRS). The team was also involved in the ROADEF's office (the national OR society) ; we also participate to the scientific committee of the GDR RO (responsible of the decision axis) and to the lead of the working group on "Algorithmic Decision and Game Theory" shared by the GDR RO and the pre-GDR IA.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Algorithmic Decision Theory. We have worked on various computational issues and challenges posed by the implementation of sophisticated decision models for individual and collective decision making. In particular, we intensively developed incremental preference elicitation methods aiming to minimize the communication cost of preference elicitation while providing efficient algorithms to determine optimal choices under incomplete preference information. This has led to various publications in multiobjective optimization, in decision making under uncertainty, and in computational social choice. A second set of results concerns the algorithmics of sequential decision making under risk. In particular, we have studied the computational complexity of determining an optimal policy in a decision tree when using very expressive models such as SSB utility and weighted expected utility (WEU). We also studied preference-based Markov decision processes. Our main contributions concern the determination of compromise solutions and independently the elicitation of reward functions. A last important set of results concerns computational social choice and algorithmic game theory. We worked on fair multiagent combinatorial optimization problems and coordination problems in agent networks (search for equilibria). In voting, we have worked on the characterization of scoring rules with distances and on the recognition of nearly single-peaked preferences and their impact on the complexity of the proportional representation problem and that of Kemeny election. Finally, we also worked on the extension of the formalism of contextual graphs for group activity modeling.

Graphical Models. Our results are fourfold : they concern the learning/elicitation of graphical models, their exploitation, the extension of existing models to cope with real-world situations, and the development of software. We proposed new Bayesian network (BN) structure learning algorithms in "non-classical" contexts, when probability distributions are non-stationary, when they include continuous variables or when there exist determinic relations among variables. We also proposed new learning algorithms for factored Markov decision processes and for imprecise probabilities. We proposed new independence testing algorithms, and new sampling procedures. We developed extensions of existing graphical models : fully object oriented Probabilistic Relation Models (PRM), in which our structured inferences outperform those in BNs by orders of magnitude, and conditional truncated densities Bayesian networks (ctdBN), which are capable of approximating almost all classical mixed probability distributions while allowing inferences as fast as those in classical BNs. We exploited these models and algorithms in several application domains, like student modeling, computer vision, cybersecurity, food science, automatic service testing. Finally, they have all been implemented in a freeware opensource C++ library called aGrUM (more than 250k lines of code) and in its python wrapper called pyAgrum.

Modeling and Optimization of Systems. Various robust multistage decision problems under uncertainty arising in connection with applications such as the Unit Commitment problem in energy management, or in the context of production/inventory planning have been investigated. The generic structure of so-called state-space representable uncertainty set has been introduced and has been shown to lend itself to efficient computation of exact optimal robust strategies for problem instances involving large number of time periods. Ongoing work is currently addressing stochastic versions of lot sizing problems and new applications in the optimal management of energy production and storage in the context of smart grids. Concerning the investigation of new models and solution methods for some large scale integer or mixed integer programming problems, linearization techniques combined with reformulations (in extended space) have been developed in particular to solve large scale graph partitioning problems and some stochastic variants of graph partitioning. In addition to this, the possibility of building more compact formulations has also been successfully explored in particular in connection with the well-known Max-Cut problem on sparse graphs, with the problem of optimizing a Choquet integral, and with the so-called Tail Assignment problem in airline operations management. Concerning the investigation of stochastic optimization problems, models involving probabilistic constraints ('chance constraints') and linear inequality systems with random (Gaussian) coefficients have been explored. Such models relate to many practically important applications such as blending problems (under uncertain composition of the ingredients) which arise in many industries (metallurgy, chemistry, petro-chemistry, food industries, etc.). New necessary and sufficient conditions for local concavity of the probability functions have been obtained both for the case of single-sided inequalities and the case of double-sided inequalities. In spite of the inherent nonconvexity of these problems, fairly sharp confidence intervals for the exact globally optimal solutions could be obtained.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	6	13	8	10	15	9	61
Ouvrages	1	4	0	0	0	0	5
Conferences	20	38	22	24	20	4	128
Logiciels	0	0	0	0	0	0	0
Brevets	0	0	0	0	0	0	0

Sélection des produits et des activités de recherche

We present 2 representative results for each of the 3 activities (ADT, GRM, MOS) developed in the team :

ADT 1 : An Incremental Elicitation Method for the Choquet Integral. The Choquet integral is one of the most elaborated decision criteria for decision making from several points of view (criteria, agents or scenarios). It is parameterized by a set function (named capacity) modeling the importance attached to any subset of points of view. The elicitation of this capacity is a challenging problem due to the complexity of the model. We have proposed the first incremental approach for the elicitation of the Choquet integral on explicit sets [**DEC-J-1**, **DEC-C-2**]. Our approach uses regret minimization to identify the most informative missing preference information and to direct an adaptive elicitation process to progressively reduce the uncertainty attached to the

capacity until a necessary optimal solution can be determined. This approach significantly reduces the elicitation burden compared to full elicitation processes. Furthermore we have combined the incremental elicitation of the capacity with a multiobjective search algorithm to incrementally determine Choquet-optimal solutions in multiobjective state space search problems [**DEC-C-3**].

ADT 2 : A New Algorithm for Robust Optimization with Interval Data. Minmax regret optimization is a popular way to cope with robustness in combinatorial optimization problems. We proposed a new lower bounding procedure for minmax regret optimization problems. This procedure is both anytime and generic, i.e. adaptable to a wide variety of problems. We take advantage of a game theoretic view of robust optimization which is very promising in our opinion. This view makes it possible to specify a double oracle algorithm (McMahan, Gordon, and Blum, 2003) to compute the lower bound in small computation times. In particular, we have successfully applied our approach to the robust shortest path problem, with significant gains in computation times compared to previous works in the literature [**DEC-J-4**].

GRM 1 : Efficient Incremental Planning and Learning with Multi-Valued Decision Diagrams. In the domain of decision theoretic planning, the factored framework (Factored Markov Decision Process, fmdp) has produced optimized algorithms using structured representations such as Decision Trees (Structured Value Iteration (svi), Structured Policy Iteration (spi)) or Algebraic Decision Diagrams (Stochastic Planning Using Decision Diagrams (spudd)). Since it may be difficult to elaborate the factored models used by these algorithms, the architecture SDYNA, which combines learning and planning algorithms using structured representations, was introduced. However, the state-of-the-art algorithms for incremental learning, for structured decision theoretic planning or for reinforcement learning require the problem to be specified only with binary variables and/or use data structures that can be improved in term of compactness. We propose a new instantiation of the SDYNA framework that uses Multi-Valued Decision Diagrams (MDD) as a more efficient data structure and describe the first planning algorithm and the first incremental learning algorithm dedicated to this data structure [**DEC-J-8**].

GRM 2 : Combinatorial Resampling Particle Filter : an Effective and Efficient Method for Articulated Object Tracking. Particle filter (PF) is a method dedicated to posterior density estimations using weighted samples whose elements are called particles. One of PF's crucial step is a resampling step in which particles are resampled to avoid degeneracy problems. We have proposed to exploit mathematical properties of articulated objects to swap conditionally independent subparts of the particles in order to generate new particle sets [**DEC-J-5**]. We have then introduced a new resampling method called Combinatorial Resampling (CR) that resamples over the particle set resulting from all the "admissible" swappings, the so-called combinatorial set. By construction, the particle sets produced by CR better represent the density to estimate over the whole state space than the original set and, therefore, CR produces higher quality samples. The combinatorial set is generally of an exponential size but, to be scalable, we show how it can be implicitly constructed and resampled from, thus resulting in both an efficient and effective resampling scheme.

MOS 1 : Exactly Solving some Robust Multistage Decision Problems under Uncertainty. A wide class of problems for which exact optimal robust strategies can be efficiently computed has been exhibited and investigated. Instances involving a fairly large number of stages (typically 50 time periods) could be solved in [**DEC-J-11**] in connection with an application to optimal energy production (Unit Commitment problem) under demand uncertainty. To the best of our knowledge, this is the first time robust optimal strategies for such number of stages can be obtained for a nontrivial decision problem. The basic idea underlying the approach consists in working out a 'compact' representation of uncertainty sets. A first proposal of compact representation in (Minoux, 2008, Algorithmic OR, 4,1 :1-18) has been generalized via the concept of state space representable uncertainty set. We have shown that this concept includes, as special cases, various types of uncertainty sets previously proposed in the literature, in particular those introduced by (Bertsimas & Sim, 2003), and those associated with discrete time finite state Markov processes. Other applications such as energy management in smart grids are currently under investigation.

MOS 2 : Quadratic 0-1 Optimization Problems, Linearization Techniques and Graph Partitioning. One of the main research fields investigated by the MOS group concerns reformulation and/or linearization techniques for solving combinatorial problems. The results obtained have given rise to 6 publications in international journals over the period. Among the significant achievements, we can mention the work carried out in connection with the PhD work of D.P. Nguyen (in close cooperation with CEA) on a stochastic version of Graph Partitioning including probabilistic constraints on the weights of the clusters [**DEC-J-12**, **DEC-J-13**] ; to the best

of our knowledge, this is the first time that this kind of problem is solved exactly for instances of significant size. Moreover, this problem has stimulated the investigation of a new variant of a linearization procedure initially proposed by Sherali & Smith (2007). This investigation has also led to compare, in terms of solution efficiency, two different ways of modeling the problem, one using a formulation based on binary second-order cone constraints, and the other one using a 0-1 quadratic formulation.

We also mention a software library :

aGrUM/pyAgrum (GRM). The team actively develops a freeware opensource C++ library called aGrUM for handling, learning and making inferences with graphical models for decision making (more than 250k lines of code). This library can be installed/run on Linux, Windows 10 and Mac OS. It is shipped with a python2/python3 wrapper called pyAgrum, which can be easily installed using pip, pip3 or conda.

Faits marquants

PUBLICATIONS

In addition to the 6 journal papers associated to the **6 results presented above** [**DEC-J-1**, **DEC-J-4**, **DEC-J-8**, **DEC-J-5**, **DEC-J-11**, **DEC-J-12**], we highlight :

• **10 IJCAI** conference papers including [**DEC-C-1**, **DEC-C-3**, **DEC-C-4**, **DEC-C-8**, **DEC-C-12**, **DEC-C-13**, **DEC-C-15**, **DEC-C-22**, **DEC-C-23**], and 4 AAAI, 4 UAI, 4 AAMAS, 6 ECAI conference papers,

• **3 best student papers** at FLAIRS'15 [**DEC-C-24**], ECAI'14 [**DEC-C-2**] and ICIP'12 [**DEC-C-5**],

• **1 best paper** at MIWAI'13 [**DEC-C-21**].

SOFTWARES

• Our library **pyAgrum** has been downloaded using conda more than 7000 times since the beginning of 2017 (statistics from https ://anaconda.org/conda-forge/pyagrum/badges).

3. Analyse SWOT

	Helpful (to achieve the objective)	Harmful (to achieve the objective)
Internal origin (team attributes)	Dual AI-OR culture + expertise in Decision Theory Synergies between the 3 activities of the team Dual competency in models and algorithms Activity between theory and practice	Reduction of the workforce (mobility, retirement) Difficulty to be simultaneously present in AI and OR research groups and projects Lack of a permanent engineer to maintain libraries of algorithms and support software developments
External origin (environment attributes)	Interactions with RO, SMA and MOCAH teams in DESIR, and other LIP6 and ISIR teams UPMC students (ANDROID) Cooperation with other labs (LAMSADE, IRIT, CRIL...) GDRI ALGODEC, GDRs RO & IA, PGMO program Growing impact of AI in industry	Scarcity of positions currently open to recruitment Gap between some challenges addressed in the international academic world and the local demand from industry Competition with other teams/labs

4. Projet scientifique à cinq ans

Concerning algorithmic decision theory, we have identified three important challenges. The first one concerns incremental preference elicitation and preference learning ; obtaining preferences remains indeed the bottleneck in many recommendation tasks. Our objectives are to propose efficient elicitation procedures based on non-linear aggregation models for interactive decision support, to produce theoretical results to bound the number of preference queries in elicitation procedures (query complexity), to design new elicitation protocols robust to local inconsistencies in answers to preference queries, and to tackle preference elicitation in ranking problems. The second direction concerns preference-based optimization with complex decision models (complexity and algorithms) in multi-objective combinatorial optimization, in multi-agent optimization and in sequential decision problems under uncertainty. We will not only work on the determination of deterministic optimal solutions but also on the determination of randomized solutions (characterized by a probability distribution over feasible solutions). Finally, a third direction concerns computational social choice. We will work on generalizations of common ranking procedures in social choice, as well as on the study of the computational problems it raises. We will also pursue our investigations on multiagent fair allocation problems.

Concerning graphical models, the team will continue working on large scale graphical models, essentially PRMs, focusing on learning and incremental structured inferences. We will also continue our work on learning non-stationary and mixed probability distributions. For this purpose, different tracks will be sought : coping with both slow distribution shiftings with strong changes (classical transition times), relaxing the classical discrete time requirements to cope with different time scales, learning non-stationary mixed distributions, learning with time sequences of different sizes (common in robotics). Our applications in cybersecurity also require better models for detecting rare or unusual events. Finally, we will start a new thread of research on learning and inference in causal models. To make all these models and algorithms available to the community, we will actively continue the development of the aGrUM/pyAgrum library.

Concerning the optimization of systems, one of the main research projects concerns some representation issues for combinatorial polyhedra related to the determination of the minimum size of a polynomial extended formulation. Of special interest will be the case of the max-cut polyhedron associated to a given sparse graph, for which encouraging preliminary results have been obtained. Another important subject for future investigation will concern the solution of some hard non convex nonlinear optimization problems arising in the context of random linear inequalities (with applications to robust optimization in random linear systems). We also mention that the solution methods previously developed for exactly solving robust multistage optimization problems constitute a natural tool for exploring the key issue of comparing (e.g. in terms of risk) the optimal strategies produced by the two main existing approaches (stochastic, robust) to decision under uncertainty, namely stochastic approaches and robust approaches. This issue does not seem to have been addressed so far on a quantitative basis in the literature and our preliminary results on a family of small sized examples seem to represent the first such approach. Future investigations will aim at confirming and extending these results by carrying out more systematic comparisons on bigger problem instances, closer to practical applications (e.g. production/inventory planning problems) ; the ultimate aim being to better characterize (e.g. in terms of risk) the domains where each of the two approaches can be expected to yield best performances.

Beside this theoretical work, we intend to develop our transfer activities. The potential applications of graphical models are multiple and already lead to various cooperations of the team with industry, as well as European projects ; a growing demand can be expected on this line. On the other hand, the transfer to industry of model-based decision support tools developed by the team is still under-exploited and should be enhanced. There are several significant contextual arguments in favour of the development of industrial applications involving intelligent model-based decision systems. Among them, let us mention the growing impact of AI in industry and the fact that recent European laws concerning automated systems possibly impacting human beings require that any decision made by a system should be possibly explained and justified [1]. However, increasing the transfer activity while preserving the quality of our academic impact is challenging. The work-force of the team has been partly impacted by the mobility of three permanent colleagues in the three last years and some other moves due to promotions or retirement are to be expected in a near future. The scarcity of positions open to recruitment in computer science these last years at UPMC did not allow us to recruit new colleagues since 2012 and this becomes now a priority for sustaining our research activity and involvement in master courses.

1. see Goodman, Bryce and Flaxman, Seth (2016), European Union regulations on algorithmic decision-making and a "right to explanation", eprint arXiv :1606.08813, 06/2016.

EQUIPE MOCAH

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

During the reorganization of the LIP6 in 2005, the MOCAH (Models and Tools in Knowledge Engineering for Human Learning) research team has been created by Jean-Marc Labat and has been part of the DESIR department. Vanda Luengo has taken the lead of the team since December 2016. The team is located on the Jussieu campus.

The MOCAH team is specialized in Technology Enhanced Learning (TEL) with an AI background (knowledge engineering, ontologies, conversational agents, data mining).

Tableau des effectifs et moyens de l'équipe

In July 2017, the team is made of 6 permanent researchers : 2 professors (V. Luengo, J-M. Labat) and 4 assistant professors (T. Carron, A. Yessad, F. Bouchet, M. Muratet), including 1 with HDR (T. Carron). Although it is not visible in the table below, the team has been heavily renewed over the past 5 years, with 4 assistant professors retiring (O. Auzende, H. Giroire, F. Le Calvez, E. Delozanne), 1 associate professor joining the team (T. Carron), 2 new assistant professors (F. Bouchet, M. Muratet) and 1 full professor (V. Luengo) recruited. 6 PhD theses have been defended over the evaluation period.

Besides we can see that the contractual activity has strongly increased during this period, without being detrimental to our regular publishing activity in the top conferences of our domain. However, publication activity in journals needs to be reinforced.

Workforces	2012	2013	2014	2015	2016	2017
Professors	7	8	5	7	6	6
Researchers	0	0	0	0	0	0
PhD students	6	7	7	10	10	8
Others	2	3	7	9	6	2
Total	15	18	19	26	22	16

Grants	2012	2013	2014	2015	2016	2017
International	0 k€	0 k€	0 k€	0 k€	0 k€	0 k€
National	58.2 k€	169.9 k€	165.9k€	98.4 k€	160.8 k€	146 k€
Industry	10 k€	0 k€	0 k€	0 k€	0 k€	0 k€
Others	0 k€	0 k€	0 k€	75 k€	200.8 k€	170.3 k€
Total	68.2 k€	169.9 k€	165.9 k€	173.4 k€	361.6 k€	316.3 k€

Politique scientifique

The MOCAH team has a high activity linked to real usages and real applications. Most of the research work involves end users in real situations, i.e. experimentations with students and/or teachers in primary schools (learning analytics dashboard for adaptation), high schools (mathematics diagnosis), university (serious games design and adaptation, recommendation in MOOCs, learning analytics from forums for student models, adaptation model for Moodle), professional training (diagnosis and decisional system for feedback using virtual

reality). Results are valued in the best conferences of our domain (Technology Enhanced Learning), several tools and software are shared with the community (Hubble project, Prog&Play), in particular through our GitHub repository (`https://github.com/Mocahteam`), or integrated into existing frameworks (Unity, Canvas). One of these results was valued by an « enveloppe Soleau ».

Besides, the team has an important involvement in the national TEL association (ATIEF), the national network granted by the ANR (Orphée), and it plays a leading role in the policy of educational innovation within UPMC but also in relation with the new Sorbonne Université (Vanda Luengo is commissioned by the vice-president to propose a new structure).

2. Produits de la recherche et activités de recherche

Bilan scientifique

The research themes of MOCAH have evolved for the last 10 years, with a reuse and development of our expertise into new themes. We have been working on authoring tools and cognitive diagnosis, using mainly symbolic approach. The core of our research is proposing models that embed human knowledge in TEL systems to improve learner models, diagnosis, feedback and the systems themselves. Knowledge is gathered (1) directly from humans (experts, learners, teachers...) using knowledge engineering techniques and/or (2) from data using educational data mining techniques.

Regarding cognitive diagnosis, 4 PhD theses were defended (2010-2015), working on designing rules for error diagnostic in math exercises, using different AI models such as Bayesian networks and, more recently, refining them using data mining techniques to propose a re-ordering of questions for a faster online identification of students' strengths and weaknesses.

Since 2008, we re-invested our expertise in the field of serious gaming with 3 PhD theses defended (2013-2015) leading to the development of various software : a framework to implement serious games, a cognitive diagnosis coupling Petri net and ontology, and an authoring tool allowing tutors to graphically redefine the scenario taking into account the context and his/her pedagogical choices.

Following the 2013 recommendation of the advisory board, we expanded research on serious gaming (4 ongoing PhD theses) and also started working on some aspects of the MOOCs (in collaboration with Univ. of Sydney and French startup Unow), extending our expertise about educational traces analysis but in the context of a large amount of traces (3 ongoing PhD theses).

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	7	3	5	3	3	1	22
Ouvrages	2	1	2	0	2	0	7
Conferences	12	17	11	19	28	17	104
Logiciels	1	0	1	0	3	1	6
Brevets	0	0	0	0	0	0	0

Sélection des produits et des activités de recherche

The selection is provided in "Annexe 4". In the first part of section I (Produits de recherche) we have selected 5 key publications which illustrate the activity of the team and are representative of the kind of ongoing research. 3 papers show our research activity on serious games (analysis of student behavior, authoring tools) and 2 papers show analysis of multimodal and heterogeneous traces in various e-learning contexts. In the second part, we present 20% of our publications in each category.

Faits marquants

Following on from the Phd of Bruno Capdevilla on "Serious Game Architecture and Design" defended at UPMC in 2013, the MOCAH team has been designing a new version of this work with a modern technology (Unity). The team has also been developing the FYFY plugin for Unity engine. Two hightlights are associated to this work : we filed a SOLEAU envelope to record the creation date of this contribution and we also use this

plugin to teach Entity-Component-System design methodology into the computer sciences Master of UPMC (ANDROIDE specialization). Making the link between research and teaching is a key point in our community.

The transfer of research into ecological teaching is also performed through the Prog&Play project. Prog&Play is a serious game, initially designed and created by Mathieu Muratet during his PhD, to motivate students to write computer programs. This serious game has been used at UPMC since 2010 and has been used by more than one thousand of students in France. This project is used as a framework to experiment the impact of serious game on motivation, learning programming, teaching practices and traces analysis.

Most of our projects (2/3 ANR grant, "Direction of numérique" grant, labex grants) work in real context and ecological situations. This scalability of our results allows us to test our approaches, models and algorithms, but it is also an opportunity to design new models from data and work with end users and educational stakeholders.

Finally, in the context of the Labex SU, one of the objective is to redesign the educational innovation in the university. The MOCAH team plays an important part in this transformation, first through Vanda Luengo's mission to propose a new structure including various educational research aspects, but also through the implication of the team in several projects related to our research policy (adaptation, learning analytics, Moocs and serious games).

3. Analyse SWOT

Strengths : MOCAH is a team doing applied research by choice but also because our financial resources have been coming mainly from projects with industrials partners and/or research partners funded by the French government. Since 2005, the team has always had at least one ongoing project, usually involving an engineer and a post-doc (5 projects on serious gaming during that period). It was crucial as in TEL it is mandatory to develop a complete application to test it in an ecological context (in classroom, not in laboratory). Another strength is that many projects are conducted in an interdisciplinary context (with colleagues specialized in cognitive psychology, didactics, medicine). For example, the latest PhD defense (July 2016) was co-supervised by E. Sander (Pr. of cognitive psychology) and J-M. Labat.

Weaknesses : Although we have a certain visibility at the international level (we are PC members of the most important TEL conferences and we have 2 PhD students totally funded by their country for 4 years [China, Thailand], and 2 partially funded PhD for long stays in Paris [Senegal]), we have very few projects at the European level.

Opportunities : With our colleagues from Universities Sorbonne Paris Cité, Paris Diderot, Cergy Pontoise, Vincennes and ENS Cachan, we decided to create an pluridisciplinary team AIDA (Approche Interdisciplinaire des Dispositifs informatisés d'Apprentissage, Interdisciplinary Approach for Interactive Learning Environments) to have a bi-mensual seminar (`http://aida.lip6.fr`). It is a great opportunity to constitute a team with a critical size with many disciplines represented (educational sciences, didactics, information sciences, cognitive psychology and computer sciences). It balances the fact that MOCAH team is rather small, which is a difficulty to maintain international and national visibility and to have the capacity to answer to project calls.

Another opportunity is the possibility to have experimental fields with numerous users in the French context with different educational level from primary schools (thanks to the new collaborations with the digital education department of the ministry of education) to the university level (thanks to the chair from Sorbonne Universités and the collaborations with the new center of innovation and pedagogy in our university).

Our research team is also heavily involved in managing and teaching in the Master 2 IMFL (Engineering and Management of E-Learning) which provides a formation to a dozen of students (mostly current or future school teachers) every year. Those students provide us with an essential network of contacts when they go work with e-learning companies for their internships, go back to their schools and/or change their career for instance to work with the ministry or board of education (rectorat).

Threats : The team is small, 6 professors (4 MCF, 2 PR, in 2018 only 1 PR), with a big activity in applications and industrial collaborations. The need for quick results stemming from industrial partnerships is not always compatible with difficulties coming from data collection in ecological situations, and can sometimes harm research results. Besides, the maintenance and the need for continuous development of the research frameworks we develop are challenging without a dedicated engineer to help in these tasks.

4. Projet scientifique à cinq ans

Research policy in MOCAH team has evolved to propose mixed approaches (symbolic and numeric) to build models to supervise and diagnose students' behaviors, knowledge, competencies and models able to propose a more adapted feedback in relation to didactical and pedagogical constraints. These models are associated to open computer environments such as simulations, serious games and virtual reality systems, i.e. environnement with higher possibilities of interaction. Other less interactive environments like MOOCs interest us to address the issues they raise about the learner assessment, feedback and adaptation scenarios from large amount of data.

Currently MOCAH is striving to experiment in real situations with large number of students and various domains and levels. The team has obtained a chair (V. Luengo) from Sorbonne Universités in 2015 to apply their research to our university context. Additionally, MOCAH participates to 3 ANR projects, and has a contract with the DNE (Direction du Numérique Educatif). All these projects are funded until 2018.

In years to come, MOCAH is planning to reinforce its research in the following themes :

— Tracking, diagnosis and feedback taking into account data heterogeneity coming from various sources (activity from simulator, serious game, LMS, MCQ, etc.) and not only actions but also perceptions (eyetracking, haptic...). We are interested in exploring the potential of new paradigms such as connected objects (IoT), mobile learning, virtual and augmented reality, and open collaborative platforms.

— Working in merging knowledge extracted from data with the ones built from humans.

— Learning analytics methods and tools to help stakeholders, in particular teachers, to adapt and personalize learning. This involves in particular the use of techniques from Natural Language Processing to analyze students' interactions with teachers and among themselves to build better knowledge models and/or propose adapted activities like students' groups.

Also, MOCAH is planning to reinforce its international research collaborations and to find a better balance between publications in higher conferences (currently high) and journal papers (currently average).

EQUIPE RO

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The Operations Research team works on combinatorial optimization problems, with many applications to scheduling, network design but also lot-sizing, operations and supply chain management for example. We approach problems with algorithmic game theory, online and approximation algorithms, fixed parameter algorithms, self-stabilizing algorithms, search heuristics, blackbox complexity, mathematical programming, Markov decision processes and stochastic optimization. Our research both aims developing the theory and designing solutions for pure problems or real life applications.

The team consists of faculty members from the hosting university UPMC as well as from the *Université Paris Nanterre*. In addition there are CNRS researchers, an Emeritus Professor, and we host a researcher from the company *Electricité de France* (EDF) who is part time in LIP6 for her research.

We have a regular flow of students (master internships or PhDs), with 6 doctoral defenses during the last 4 years.

Tableau des effectifs et moyens de l'équipe

PhD Students co-supervised with another university are counted half.

Workforces	2012	2013	2014	2015	2016	2017
Professors	8	8	9	9	9	9
Researchers	3	3	4	4	4	4
PhD students	7.5	7	3.5	1	2	3.5
Others			1	1		
Total	18.5	18	17.5	15	15	16.5

Grants	2012	2013	2014	2015	2016	2017
International						
National	35	36	81	41	20	14
Industry	52	24	10	4	20	20
Others						
Total	87	60	91	45	40	34

Politique scientifique

The operations research team studies various combinatorial optimization problems with a particular interest in scheduling and graph problems, which contributes to the *Science des données, intelligence et optimisation* research axis of the LIP6. Our global approach facing an optimization problem consists in developing a mathematical model, searching for its combinatorial complexity, finding good properties of its optimal solutions and deriving an exact or approximation algorithm. Performance is either measured experimentally or by providing worst case guarantees.

What is a team, but their members ? Here they are in alphabetic order : Spyros Angelopoulos (CR, chargé de recherche) works on online algorithms and search games. Evripidis Bampis (PR, professeur) works on algorithmic game theory and approximation algorithms. Philippe Chrétienne (Professeur émèrite) works on scheduling

and network design. François Delbot (MCF, maître de de conférences) studies self-stabilizing algorithms. Carola Doerr (CR) works on search heuristics and blackbox complexity. Olivier Dubois (CR) studies random instances of constraint satisfaction problems. Christoph Dürr (DR, directeur de recherche) works on online and approximation algorithms. Bruno Escoffier (PR) designs fixed parameter and approximation algorithms. Pierre Fouilhoux (MCF) works in mathematical programming for combinatorial optimization problems. Claire Hanen (PR) works on resource constrained scheduling. Emmanuel Hyon (MCF) studies Markov decision processes and stochastic optimization. Safia Kedad-Sidhoum (MCF) studies scheduling, lot-sizing, operations and supply chain management. Fanny Pascual (MCF) works in algorithmic game theory and approximation algorithms.

Our team has made contributions in particular in scheduling problems that appear in telecommunication networks, in industrial processes, or in the speed scaling setting. We also studied algorithm game theoretical aspects of scheduling, and search heuristics or search problems.

We are involved in international research communities, which is witnessed by conference organizations (STACS, SEA, WAOA, GRASTA) and steering committees (ISCO, STACS, WAOA and IWLS, which we co-founded), as well as the organization of several focused workshops (Women@GECCO, spring school on algorithmic game theory, etc.). The team has formal bilateral exchange programs with the Netherlands (Nikhil Bansal), Germany (Nicole Megow), Greece (Ioannis Milis), Poland (Krzysztof Rzadca), Argentina (Eugenio Della Vecchia) and Chile (Fernando Ordonez), but collaborates as well for example with Thomas Lidbetter (UK), Chien-Chung Huang and Bengt Nilson (Sweden), Dimitrios Fotakis (Greece), Halil Sen and Kerem Bulbul (Turkey), Alexander Kononov (Russia), and Maxim Sviridenko, Marek Chrobak or George Rouskas (USA).

Among our national activities, we could mention the *CNRS Groupement de Recherche en Recherche Opérationnelle (GdR-RO)* which was created by Philippe Chrétienne, managed by him until 2012, and where Pierre Fouilhoux and Emmanuel Hyon are part of the management committee. We animated several working groups within this GdR-RO, namely POC, P2LS, COSMOS and AGAPE.

During the last 4 years we obtained financial supports from various sources. Among others we had a Google Focused Research Awards, a research project with *Electricité de France (EDF)*, a contract with the *Direction Générale de l'Armement (DGA)* on optimization problems arising in radars. Several grants from the *Agence Nationale pour la Recherche (ANR)* (Netoc, Marmotte, Cocorico, Coca and Oata) and *The Gaspard Monge Program for Optimization and operations research (PGMO)* enabled us to collaborate nationwide with universities in the Paris area, from Grenoble, Saint-Etienne or Toulouse.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Our main scientific production concerns combinatorial optimization. Some of the results are application driven, while others are motivated by the employed methods.

Design and analysis of algorithms whose complexities are beyond polynomial time has become a major research axis, in particular the field of parameterized complexity (fixed parameter algorithms) where we express complexity as a function of several parameters (not only the size of the instance). We studied several classical NP-hard problems in this line of research, with a special focus on approximation algorithms, exhibiting strong lower bounds about subexponential time approximation algorithms.

The area of *online algorithms* deals with the paradigm of the input arriving in form of requests, and each request must be served by action by the algorithm so to maintain a feasible solution for the part of the input seen so far, while optimizing some objective functions. We studied clustering and scheduling problems in this setting. The performance of an online algorithm is measured by the so called *competitive ratio*, comparing in a worst case scenario the solution produced by the algorithm with the optimal *offline* solution. This ratio measures the price of not knowing the future requests. We loose this setting by allowing the algorithm to obtain some information (*advice*) about the whole instance. Then the competitive ratio is a function on the amount of provided advice, and we studied this dependency for the classical bipartite matching problem.

One of the main beyond-the-worst-case-models is multi stage stochastic optimization. Here the input arrives in stages, and the algorithm knows the distribution from which the input is drawn. These kinds of problems cover a large application field such as networking (admission control, power allocation in wireless networks), cloud (energy saving, virtual machine management) or inventory control. One of the innovative aspects is the

consideration of *impatience* in the system. We mainly address these problems with the stochastic dynamic control framework by optimizing a given objective function in case where the behavior of the system is driven by random events and by controls. Our contribution to this field is twofold. First we develop a package for Markov Decision Processes within the ANR project Marmote. Second we try to discover optimality conditions of policies that owns a particular structure (eg a threshold policy). We also study game theoretical aspects of these models.

Randomized search heuristics (RSH), are general-purpose algorithms and are *the* most commonly applied solvers in industrial and academic optimization. Understanding RSH with mathematical precision is a fundamental challenge in computer science. On the one hand, it helps to identify how certain algorithmic choices influence the performance of a heuristic. This is very crucial, in particular, for *large-scale optimization,* where the order by which the optimization time grows determines which problem sizes can or cannot be solved by a given approach. But theory of RSH is much more than "just" understanding optimization time distributions. More importantly, it offers a complementary view on experimental evaluations of RSH and plays an important role in identifying the *working principles* of these algorithms. This knowledge is essential for the development of new heuristic solutions, witnessed, for example, by the series of recent works on *self-adjusting* parameter choices in RSH. Last but not least, the theory of RSH also offers a plethora of *beautiful algorithmic and probabilistic problems,* which often inspire the development of new mathematical tools and give alternative views on classical problems in theoretical computer science (TCS).

Designing novel models and approaches for sustainable supply chain management is an emerging research area. We studied the way carbon emission constraints are integrated and how to handle their induced complexity within supply chain planning models. We proposed and studied different types of carbon emission constraints : periodic, cumulative, global and rolling carbon emission constraints. We showed that the multi-sourcing uncapacitated lot-sizing problem with periodic carbon emission can be solved optimally using a dynamic programming algorithm. We also proved that, when considering one of the other three constraints, the ULS problem becomes NP-hard.

We wanted to go a step further by analyzing an efficient way of adopting the general concept of the circular economy by designing the supply chain as "circular" by eco-designing products and considering remanufacturing, recycling and reuse of products, and the use of renewable energy. We studied planning decisions of re-manufacturing activities and considered uncertainties on both the returns quantity and quality and developed new stochastic programming based approaches. Modern platform architectures are built with a hybrid structure constituted of multi-core CPUs coupled with several GPU accelerators. They generate challenging scheduling problems that we studied. We proposed several algorithms with constant approximation ratios in the case of independent tasks with a reasonable time complexity that allows a practical applicability. Some extensions with dependent tasks as well as moldable tasks have been considered.

We are interested by sharing resources among several agents, where a resource can be a machine or a set of machines — for scheduling problems — or a set of items — for matching problems. Our aim is to obtain good solutions for the social welfare, despite the fact that there are agents with conflicting interests. We gave truthful algorithms, in which the agents cannot manipulate the system by lying about their private data, and we gave mechanisms where the agents have incentive to cooperate in order to reach a solution which is good for all of them. We also studied data center resource management, where we aim to assign different tasks to the machines of a data center in order to maximize the utilitarian or the egalitarian objective (each task being owned by an agent). Here the difficulty is that the tasks have different types which reflect their nature (CPU-intensive tasks or bandwidth intensive tasks for example).

We have worked on algorithmic issues on search theory. In these types of problems, a mobile searcher must locate a target who hides at some unknown position of the environment (modeled, e.g., as a graph), as efficiently as possible. Such problems arise very often in operations research, theoretical computer science, and artificial intelligence. More precisely, we worked on the following topics : i) Measures and algorithms for *expanding search* over a graph : here, we considered the setting in which the searcher incurs cost only during the first time it traverses the edge of a graph (this setting can model, for instance, drilling for oil in multiple locations). We gave algorithms and hardness results for a variety of graphs such as stars, trees and general undirected graphs, using the *competitive ratio* as measure ; ii) *Infinite linear programming* and duality in searching : here, we demonstrated a critical error in an earlier paper on this topic, and showed how to properly use infinite linear programs and duality so as to obtain optimal search algorithms in the setting in which the searcher incurs a turn

cost every time it changes directions ; and iii) *Multiprocessor* search and scheduling problems with setup cost : in this work, we used linear programming techniques in order to provide near-optimal algorithms for problems that involve multiple searchers. The results generalize various previous work that focused on a single searcher, and apply also on more general scheduling problems in which every execution of a task incurs a fixed setup cost.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	14	13	17	16	17	18	95
Ouvrages					1		1
Conferences	18	20	13	16	20	3	90
Logiciels					2		2
Brevets							

This table is not covering 49 communications in conferences without proceedings, such as the main french annual conference in operation research ROADEF, (*le congrès annuel de la société Française de Recherche Opérationnelle et d'Aide à la Décision*) and similar venues.

Sélection des produits et des activités de recherche

Faits marquants

In *robust optimization* one has to produce a solution to some instance, that can later be adapted with little cost to changes in the instance. This contrasts with the classical *input-computation-output* paradigm. Quite some different models have been considered in this context. For example in [**RO-A16**] we studied a classical scheduling problem, where we are given jobs each with some processing time and a precedence relation between them (if $i \prec j$ then job j cannot start before the completion of job i). A schedule consists of starting times such that job executions do not overlap and satisfy the precedence relation. For the two-stage robust variant of this problem each processing time comes with an interval, and the second stage instance can change within these intervals. Suppose we are given a *baseline* schedule for the first stage instance together with the second stage instance. processing times for the jobs. The goal is to produce a feasible schedule for the second instance maximizing similarity with the baseline schedule (the *anchorage level*). We provided a polynomial time algorithm for this problem. However the main challenge in this context is to come up with a good *baseline* schedule, that allows maximum anchorage level in the worst case. We showed that this problem is NP-hard when the instances have in addition time windows constraints (the starting time of each job must belong to a given interval).

Another aspect of robust optimization that was addressed in this context is the influence of information about future instances on the quality of the solution one can produce. In [**RO-F13**] we considered the classical maximum cardinality matching problem for a bi-partite graph $G(U, V, E)$. Here the U-vertices arrive on-line together with their adjacent edges, and ask to be possibly matched to some V-vertex. Decisions are irrevocable. It was known since 1990 that an on-line algorithm can approach within factor $1/2$ the optimal solution, which improves to $1 - 1/e$ if the algorithm has access to a source of random bits. Both results are tight. A sequence of papers since 2011 analyzed the achievable ratio when the algorithm has access in beforehand to some information (*advice*) about the whole instance. One of our contributions was to show that a linear number of advice bits suffices to approach the optimal solution arbitrarily close (which needs $O(n \log n)$ bits to be completely described).

Energy efficiency has become a crucial issue in Computer Science. New hardware and system-based approaches are explored for saving energy in portable battery-operated devices, personal computers, or large server farms. These last years, there is also an increasing interest in the development of algorithmic techniques for finding tradeoff-solutions between energy consumption and performance. On the scheduling level one can influence energy consumption with two main mechanisms. First the processor speed can be varied during execution time. When a processor runs at speed s during one time unit, then consumed energy (i.e., the power) is $f(s)$ with f being a non-decreasing convex function of the speed, typically $\gamma + s^\alpha$ some physical constants $2 \leq \alpha \leq 3$ and $\gamma > 0$. Second the processor can simply be turned off during idle periods, resulting in no energy consumption at all, except during the shutdown and reboot phases. We studied scheduling problems in the speed

scaling model, the power-down model and in the combination of both. In particular we conducted a systematic study of these models in the most important scheduling models, such as single or multiple processors, allowing or not preemption of jobs, instances with release times on jobs, with deadlines, with structural assumptions on the release time-deadline intervals, with precedence relations arising from the map-reduce framework and so on. We liked in this work the impact it could eventually have in the environmental footprint of the information technology, but we were also attracted by the non-linear aspects these models introduce. To handle these aspects new dynamic programming techniques needed to be developed, but we also used configuration linear programs, or reductions to convex cost flows. Moreover we studied game theoretical aspects of the speed scaling model, where players want their jobs to be executed as quickly as possible on a central server and consumed energy has the be charged to the players. The goal is to find a compromise between overall energy consumption, and to ensure standard desirable properties of games, such as truthfulness and small price of anarchy. In overall 4 PhD thesis were conducting on these problems, and produced scientific results published in 9 journal papers and 17 conference communications, such as [**RO-A11**, **RO-A12**, **RO-D2**, **RO-D4**].

Graphics Processing Units (GPU) have hi-performances on some computational tasks (which are not necessarily related to graphics). These hybrid CPU+GPU computer architectures create the need for new scheduling algorithms. We considered the central problem of assigning n given independent jobs of various length to the processing units so to minimize the latest completion time among the jobs. What makes the problem attractive is that we first need to bi-partition jobs (CPU and GPU) and then in each class to assign jobs to the cores (which is a well understood sub-problem). We achieved an approximation ratio of $4/3 + 1/3k + \epsilon$ for k GPU's which drops to $4/3 + \epsilon$ for $k = 1$. This theoretical result was completed by an experimental analysis for different kernels to assess the computational efficiency of the proposed methods. The results obtained the best paper award in the conference HeteroPar [**RO-A18**, **RO-F2**, **RO-D3**].

The paper [**RO-F10**] has received a best paper award at the 2016 edition of ACM GECCO, the prime conference on Genetic and Evolutionary Computation Conference, for its contribution to understanding the limits of randomized optimization heuristics. *Black-box complexity* is the most important performance measure for *black-box algorithms*; an important class of algorithms that optimize without having access to the problem at hand other than by sampling (deterministically or at random) potential solution candidates and receiving the function values for these. The sampling is adaptive and the performance of a black-box algorithm on a problem class is measured by the number of function evaluations that the algorithm has to perform (in expectation) until it queries, for any instance, an optimal solution for the first time. A result by Afshai, Agrawal, Doerr, Doerr, Green Larsen, and Mehlhorn [LNCS 8066, 2013] shows that the black-box complexity of LEADINGONES, a classical benchmark problem in the theory of evolutionary computation, is $\Theta(n \log \log n)$, which is substantially lower than the $\Omega(n^2)$ runtime of classical evolutionary algorithms. Doerr and Lengler prove that any algorithm achieving a $o(n^2)$ complexity has to use memory ("population") or non-elitist selection strategies (i.e., strategies to chose which previously evaluated solution candidates influence the sampling of new candidates), thus shedding light on the importance of these two commonly found but theoretically not well-understood features of evolutionary algorithms.

In addition we would like to mention the journal publication [**RO-A1**], where we solved a long standing open problem in the area of *discrete tomography*. Given a 2-dimensional grid with colored cells using k colors, the *projections* are numbers that tell for each row and column the number of its cells for each color. The *inverse* problem consists of producing a grid satisfying given *projection* constraints. It was know since the 50's that for $k = 2$ colors the problem can be solved in polynomial time, and since the late 90's that for $k = 4$ colors the problem is NP-hard. Our contribution was to settle the remaining case $k = 3$.

3. Analyse SWOT

Strength The operations research team has a long history on research on scheduling and other combinatorial optimization problems. It attracted 4 CNRS researchers which arrived respectively in 2010, 2011, 2013 and 2017 as well as a professor in 2013. This permitted is to open and enforce our research areas to algorithmic game theory, approximation and online algorithms, and fixed parameter algorithms. We have a good production of journal publications, and a strong network of international collaborations. Also we have good collaborations with other teams from LIP6, through internal projects financed by the lab or by co-supervision of PhD students.

Weakness Currently, despite some tentative, we don't have any European grant, but we try again this year.

Also besides some mainly best paper awards, we have not obtained prestigious national prices, such as membership at the *Institut universitaire de France*. In addition we have not completely exploited the existing potential for collaboration between team members.

Opportunities Our team has the possibility to make fruitful collaborations with industry in France and abroad, in particular through Cifre PhD grants.

Threats The composition of the team does not contain enough professors or assistant professors with the habilitation (HdR). The risk is that we are not enough present in the master programs, even though we teach in the masters Android, STL, MPRI and in the *Ecole Centrale*. Also in 2017 we were facing difficulties to raise the necessary financial support for our growing team.

4. Projet scientifique à cinq ans

Our scientific goal could be summarized as developing the theory needed to solve efficiently practical applications.

In the coming years we plan to keep the pace of our current scientific research, and explore new domains in combinatorial optimization, invent new tools while pushing current techniques to their limits. As for particular application areas we could mention new circular economic models or new urban transportation systems that need to be studied and call for new developments in stochastic optimization and mechanism design. In general our research will focus on problems beyond the classical algorithmic setting, such as high performance computing, multistage robust optimization or processing dynamic input under specific probabilistic models.

Our team will keep attracting industrial partners, and study practical and theoretical aspects of their problems. In addition we plan to use our international network to obtain a European grant on some hot area in combinatorial optimization, similar to the COST project on *Improving Applicability of Nature-Inspired Optimization* by Joining Theory and Practice that we are vice-chair of.

Several models have been considered and studied (in particular in our team) for problems which evolve with time (on-line algorithms, re-optimization,...). Beside computational complexity and quality of solutions, a major aspect of dynamic problems is to build or maintain solutions that do not change too much over time. We plan to incorporate this stability requirement in over-time optimization models and problems, leading to hopefully interesting tradeoffs between quality and stability of solutions.

In classical scheduling problems we have to minimize some given objective value. However when scheduling public events that make use of a shared facility, different individuals express different objective values they want to have optimized. In the underlying scheduling problem, machines represent shared facilities and resources, while jobs represent public events. This creates the need to extend tools from computational social choice in order to produce schedules that would best possible for the society, namely making every individual happy and minimizing a social cost which aggregates the individual preferences.

Mid- and short-term planning of re-manufacturing activities in reverse supply chains, are source of new interesting mathematical optimization problems. One of the main challenges is the high level of uncertainty in the input data needed to make these planning decisions. We will seek to explicitly take into account in our models the uncertain nature of the problem input data. For this purpose we intend to extend our first contributions to more complex and realistic problems and develop new stochastic programming based approaches. We aim also at developing lot-sizing problems in the context of energy management where non-reversible energy sources are used to fulfill a discrete demand over a planning horizon. In addition, reversible sources (such as battery and super-capacitor) can be used to store and/or to supply energy assuming a limited capacity. A careful management of the energy storage is required to optimize the total production cost. The non-reversible sources are characterized by an efficiency function allowing to get the amount of usable energy for a given cost. The inverse of this function is used in lot-sizing models to get the cost related to the produced amount of energy over the horizon. Efficiency aspects have to be considered for the reversible sources. Losses can also be assumed when carrying energy units from a period to another. Solving this new class of lot-sizing problems is a challenging and promising area from both theoretical and practical perspectives.

We want to study deterministic scheduling problems with non-regular performance measures and machine unavailability constraints. Following our first results for the single-machine total earliness/tardiness scheduling problem around an unrestrictive common due date with machine unavailability constraints, we aim at finding some structural properties, complexity results, and algorithms for job-dependent penalties.

Apart from developing practice-driven and -relevant algorithmic solutions, we will also put a stronger focus on performance measures that are practice- and/or implementation-driven. In particular in the domain of heuristic optimization there is an apparent disconnect between theoretical bounds and implementation-based performance.

EQUIPE SMA

DOSSIER D'AUTOÉVALUATION

1. Presentation of the team

0.1 structure of the team

The group is headed by Amal El Fallah Seghrouchni and is composed of 10 faculty members.

— 6 Associate professors : Maha Abdallah, Aurélie Beynier, Vincent Corruble, Zahia Guessoum (HDR), Cédric Herpson, Jean-Daniel Kant (HDR).
— 3 Professors : Philippe Codognet, Amal El Fallah Seghrouchni and Nicolas Maudet.
— 1 Senior researcher : Jean-Pierre Briot (DR - CNRS).
— The group hosts an associated Senior Researcher from Sony CSL : François Pachet and an associated Assistant Professor from IPSA : Assia Belbachir.

0.2 international background

Introduction

The research group was created in January 2006. We recruited Aurélie Beynier in 2008, Nicolas Maudet in 2011, Maha Abdallah and Cédric Herpson in 2015. Jean-Pierre Briot fully returned to the group in 2015 after having been Director of the CNRS office in Brazil, and Philippe Codognet joined us in 2017 after having been director of the Japanese-French Laboratory for Informatics in Tokyo (UMI 3527) and Director of the CNRS office in Japan. François Pachet has been associated to the group since the starting of his ERC Flow Machines Project in 2012. Assia Belbachir is associated with the group since 2017.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	6	6	6	7	8	9
Researchers	0	0	0	1	1	1
PhD students	15	20	20	22	12	7
Others	4	4	3	3	4	4
Total	25	30	29	30	25	21

Grants	2012	2013	2014	2015	2016	2017
International	3	3	2	1	2	2
National	5	5	4	4	2	4
Industry	8	8	11	8	5	3
Others						
Total	16	16	17	13	9	9

Politique scientifique

The group is interested in cognitive agents as a paradigm for designing and programming complex systems to bring intelligibility to the designers. Activities are organized around three axes : (1) coordination models, (2) simulation and virtual agents and (3) agent's architectures and languages. SMA group develops multi-agent models and / or simulations in which agents are autonomous (e.g to decide, plan , learn, etc.), capable of action

and interaction with other agents (e.g. to coordinate, negotiate, etc.) and the environment (e.g. to perceive and respond).

SMA group develops complementary scientific aspects such as :

— Knowledge production and involvement in research projects : publications in major conferences (AAMAS, IJCAI, AAAI...) and journals (JAAMAS, IJAOSE,Journal of Ambient Intelligence and Humanized Computing), books, etc. Coordination of an ANR (AMANDE, Advanced Multilateral Argumentation for DEliberation), and members of several others : LARDONS (Learning and Reasoning for Deciding Optimally using Numerical and Symbolic information), R-Discover (Réseaux de robots mobiles-Couverture décentralisée de l'espace basée vision omnidirectionnelle), CoCoRiCo-CoDeC (computational social choice). Coordination of one FP7 project (TARDIS), member of an ESF-COST action (COMSOC, Computational Social Choice).

— Knowledge transfer in the context of industrial collaborations. Indeed, our work has been applied in industrial contexts for the supervision of distributed production process (Dem@t-Factory project) and for the management of smart-grids, in the TerraDynamica project that aims at developing technologies enabling the animation of realistic virtual cities by simulating the presence and activities of inhabitants and vehicles going about at their regular activities but also able to react to events and to adapt to the evolution of their environment in a credible manner. Moreover, the RCSM project aims at building a platform to support the negotiations among the different actors of a financial supply chain. The group has also been involved in 12 Ph.D Thesis in collaboration with industrial partners (i.e. CIFRE) such as Thales, IBM, Orange, Kyriba, etc. We have also developed 2 PRI (Projet de Recherche et d'Innovation) with Thales Airborne Systems : the first PRI (GFAI) is about unforseen faults within critical software while the second (RAMSES) concerns the architecture of the next generation of UAVs (multi-functions multi-sensors platforms).

— Support of MAS community at national and international levels. Members of SMA group are involved in the major events dedicated to A.I. and MAS such as AAMAS, EUMAS, ECAI, IJCAI, AAAI, etc. SMA group is also involved in the main instances that manage A.I. and MAS research. At the international level, Amal El Fallah Seghrouchni is an elected member of IFAAMAS board and EUMAS board. Nicolas Maudet is involved in the pré-GdR IA (CNRS) and he was a member of the board of the national Association of A.I. (AFIA).

2. Produits de la recherche et activités de recherche

Bilan scientifique

Over the last five years, the size of SMA group has increased significantly. SMA recruited Cédric Herpson as MC, welcomed Maha Abdallah as MC of the LIP6 and Philippe Codognet as UPMC Professor. The group was also reinforced by the return of Jean-Pierre Briot DR-CNRS. In addition, the team has welcomed, as associated member of the group, a senior researcher from Sony, François Pachet for the duration of his ERC as well as Assia Belbachir who is MC at IPSA. The arrival of these colleagues reinforced the existing themes in the team and also helped to create a new one around the creativity and the generation of content. The team has strengthened its scientific involvement in the national and international communities in the field of multi-agent systems by participating to the strategic scientific forums, to various and varied scientific committees. Indeed, the members of SMA group played various roles ranging from General Chair(s) to PC members of the major conferences of our domain. In addition, SMA organized some major events in MAS field such as AAMAS 2014 which attracted more than 800 participants. The group also reinforced his international collaborations mainly with Honiden'lab at the National Institute of Informatics (Tokyo - Japan), AI-MAS Laboratory (Artificial Intelligence and Multi-Agent Systems) at the University Politehnica of Bucarest (Romania), the University of Sao Paulo (Brazil) and PUC- Rio de Janeiro (Brazil). The scientific production was also of a very good level since the team published 38 articles in major journals of the field; 19 Books, and 117 papers in the leading international conferences. It should be noted that the group has received 3 awards for "best paper" for this period, including one in the prestigious AAMAS conference. At the contractual level, the team has carried out very fruitful collaborations with the main industrial partners such as IBM, Thales, Orange, Airbus. Thus, the team members participated in numerous large-scale projects (such as TerraDynamica), developped 2 projects of research and innovation (PRI) with Thales Airborne Systems, advised 12 CIFRE (already defended) and

participated to several ANR projects. These collaborations are intended to be intensified in particular within the framework of the Institut Tremplin Carnot Interfaces affiliated with UPMC, the INSERM and the CNRS.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	8	2	9	9	6	4	38
Ouvrages	2	1	5	2	5	4	19
Conferences	23	24	28	20	17	5	117
Logiciels							
Brevets							

Sélection des produits et des activités de recherche

Faits marquants

SMA group develops complementary scientific aspects such as :

— Knowledge production and involvement in research projects : publications in major conferences (AAMAS, IJCAI, AAAI...) and journals (JAAMAS, IJAOSE, Ambient Intelligence and Humanized Computing), books, etc. Coordination of an ANR (AMANDE, Advanced Multilateral Argumentation for DEliberation), and members of several others : LARDONS (Learning and Reasoning for Deciding Optimally using Numerical and Symbolic information), R-Discover (Réseaux de robots mobiles- Couverture décentralisée de l'espace basée vision omnidirectionnelle), CoCoRiCo-CoDeC (). Coordination of one FP7 project (TARDIS), member of an ESF-COST action (COMSOC, Computational Social Choice).

— Knowledge transfer in the context of industrial collaborations. Indeed, our work has been applied in industrial contexts for the supervision of distributed production process (Dem@t-Factory project) and for the management of smart-grids, in the TerraDynamica project that aims at developing technologies enabling the animation of realistic virtual cities by simulating the presence and activities of inhabitants and vehicles going about at their regular activities but also able to react to events and to adapt to the evolution of their environment in a credible manner. Moreover, the RCSM project aims at building a platform to support the negotiations among the different actors of a financial supply chain. The group has also been involved in 12 Ph.D Thesis in collaboration with industrial partners (i.e. CIFRE) such as Thales, IBM, Orange, Kyriba, etc. We have also developed 2 PRI (Projet de Recherche et d'Innovation) with Thales Airborne Systems : the first PRI (GFAI) is about unforseen faults within critical software while the second (RAMSES) concerns the architecture of the next generation of UAVs (multi-functions multi-sensors platforms).

— Support of MAS community at national and international levels. Members of MAS group are involved in the major events dedicated to A.I. and MAS both as Track or Area Chair(s) or Local chair (AAMAS) ; SPC (IJCAI, ECAI, PRIMA) PC-Cahir and Genral chair (IDC) or as elected members(IFAAMAS Board, EUMAS board, AFIA) or pré-GdR IA (CNRS).

3. Analyse SWOT

SMA group has expert knowledge in the domains of planning, coordination, learning and simulation. The group develops complete approaches ranging from theory to prototype (TRL4-5). The group has also a great experience of collaboration with industrial partners in these domains. The group is involved in the most important editorial boards of the relevant international journals and conferences as member of IFAAMAS and EURAMAS boards. However, the group needs more technical support and could pay more attention to potential opportunities for patents. This weak point can be overcome thanks to the involvement of the LIP6 and the SMA group in the Tremplin Carnot Interfaces, one of whose objectives is to help its laboratories in this direction.

The group is interested in cognitive Multi-Agent Systems. In the next 5 years, the group intends to develop a new axis called "Creativity and Content Generation" and to follow the existing ones of research in the following directions.

Axis : Multi-agent Simulation MAS group has been particularly active in the growing field of multiagent simulation and we intend to pursue our effort in three directions :

- **Participation** : Here we seek to explore the design of participatory simulation, where human users interact with the simulation through an avatar (agent) he/she controls. The agent will embed a cognitive model of the human and we will design specific learning algorithms to estimate the parameters of the cognitive model via the observation of the user's behaviors during the simulation.
- **Credible virtual humans** : Several projects in our group focus on the challenging task of modelling credible human behaviors, whereas most of existing works are devoted to more normative behaviors. For instance, we plan to pursue our work on affective agents, to enhance the realism of our social or economic simulations. To do so, we will continue to take inspirations from other relevant fields (e.g. psychology, neuroscience, economics, social sciences,...).
- **Multi-scale**. In many cases, a simulation needs to accommodate several viewpoints (levels, contexts...) either simultaneously or sequentially. Thus, agents can represent group of individuals of various sizes at different times, thus allowing for better performance, more meaningful interactions with users and other potential benefits. Therefore we will further develop our multi-scale simulation models with an emphasis on dynamic models.

Axis : Coordination, distributed problem solving The MAS group will continue its research work on coordination and distributed problem solving. Especially, we plan to concentrate our efforts in the following domains :

- **Distributed Planning for Coordination** : We will pursue our work on distributed planning algorithms to develop efficient methods able to scale to large size of problems and to deal with non-standard settings such as non-stationary or open environments. Part of this work will be applied to the multi-agent patrolling problem. Another part will be applied in ambient intelligence domain where spatio-temporal constraints as well as context-awerness are important issues.
- **Computational Social Choice** : This recent field studies the interplay of collective decision-making (voting, resource fair division, etc.) and computer science. We will carry on our work in this area, focusing in particular on distributed resource allocation problems, that is, fully distributed methods lacking central control where each agent is initially endowed with a set of resources and agents exchange resources locally. We will investigate to what extent the properties of the problems (for instance the model of preferences or the topology of interactions) could give some guarantees (fairness, efficiency) on the final allocation.
- **Assistance for Negotiation** : We will continue our work on the design of expert artificial assistant agents to help humans to decide and negotiate the use of shared resources (socio-eco-systems), based respectively on viability analysis and argumentation. We will further investigate the relations and dependencies between argumentation systems and negotiation protocols.
- **Argumentation and Deliberation** : Formal argumentation deploys models allowing to reason on the basis of reasons pro or con some facts, either individually or collectively (*e.g.* in a debate). We want to explore more deeply strategic argumentation allowing an agent to plan her moves in a debate or in a persuasion dialogue. We will especially address the issues of modeling (model of arguments, strategy formalization,...) and strategy computation (moves planning, learning of the opponent behaviors, on-line adaptation,...).

Axis : Agent's architectures and languages An agent architecture organizes the reasoning and computational cycle of an agent and an agent programming language offers an interface to specify the activities of an agent. The MAS group will continue to design and experiment with these two dual and complementary dimensions. More specially, in the upcoming years, we will pursue our efforts among the following directions :

- **Massively parallel computation** : We consider dedicated agent-based architectures for cooperative problem solving for combinatorial search and optimisation applications. We will pursue our work on agent-based architectures to deal with the specificities of the (massively) parallel infrastructures required to tackle this class of combinatorial problems.
- **Long-Term Autonomy** : We seek to design and implement on-board cognitive architectures allowing agents to autonomously adapt their behaviours in the real-world on a long term basis (months, years).

We will carry-on our work on robust decision-making to develop *anytime* approaches, therefore allowing for better performance in environments where optimality is not reacheable. This will lead us to further investigate the influence of time-dimension on knowledge representation and thus on the agent architecture.

— **Context-sensitivity** : We will continue to explore extensions of the standard BDI (Belief-Desire-Intention) architecture and associated agent's languages to incorporate concepts such as : arguments, for collaborative applications (to support negotiation and explanation) ; contexts and spatio-temporal dimension for ambient intelligence (to support adaptation and combination of plans).

Axis : Creativity and Content Generation We are planing to work in the domain of contents generation, and in particular in computer-based music. One topic is concerned with the use of deep learning techniques to help human users at generating musical content extrapolated from a corpus of musical examples (style). One of the scientific and technical challenge is the capacity of user global control, e.g., via controlled sampling or via reward-control (reinforcement learning) strategies, as well as the design of intelligent (agent-based) control assistance/interfaces. This work is being developed within the Flow Machines ERC Project of François Pachet at Sony CSL-Paris and UPMC/LIP6. Another one is concerned with a general architecture for sound agents, where one associates sonic events to agent behaviors and thus program music pieces as multi-agent interactions. We envision collaborations with sound designers and research groups at IRCAM for developing these ideas.

Application domains In addition to the different application contexts mentioned above, three sectors appear to be particularly promising and rich in opportunities for the coming years :

— **Smart-cities** : Smart-cities are the convergence between connected cities and smartgrids. From IoT to distributed energy resources combined with demand response, plug-in electric vehicles to active management of distribution networks and traffic regulation in intelligent transportation systems, the diversity of actors, platforms and new technologies required for the advent of smart-cities call for more sophisticated solutions. Growing investments to deal with this complexity offers opportunities for the various distributed planning, multi-level simulation and decentralized decision making solutions that we develop.

— **Autonomous Aerial and Ground Vehicles** : Within either a civil or a military context, the use of Autonomous Ground and Aerial Vehicles (UGV and UAV) is intended to become the standard. While in the civilian context the focus is mainly on the improvement of one autonomous vehicle's behaviour, military requirements led to address the management and control of fleets. In both situations, the autonomy's degree of a vehicle, and its coordination with its surrounding are key points where funds are currently channeled and where our group can make significant contributions.

— **Video Games** : The continuing growth of the gaming industry is accompanied by an evolution, both in term of uses and platforms that induces a complexification of the environments in which players evolve (open, persistent, hosted on distant and distributed infrastructures). These evolution lead to a growing need for : Automatic generation of high value-added content (e.g Non-Player Character Behaviors consistent with the universe and the players interactions' over time) ;Leveraging current distributed technologies for a highly scalable game infrastructure while providing optimal quality of experience for the user/player (e.g., cloud, fog, edge, and P2P computing). This opens up the possibility for the group to express its knowledge and expertise on both dimensions.

To these various elements we could add an extra dimension, transverse, because its an integral part of the different fields and application mentioned : Collective behaviours, social dynamics and explainability.

DEPARTEMENT NETSYS

1. Présentation du département

The Networks and Systems department gathers LIP6 research activities related to networks, systems, and distributed systems. We analyze and design solutions for constructing and managing networks, systems, and distributed systems of the future. Those themes are represented by five complementary teams : **NPA** and **Phare** focus on networks, **REGAL** and **Whisper** dedicate their research activity to Systems, while **Almasty**'s focal point is communication security and cryptography. REGAL and Whisper teams are joint with Inria Paris.

The department carries out intensive scientific activity, demonstrated by a large share of publications in top journals and conferences, and implication in numerous research projects (at the European, national, and local levels, also including participation to Labex SMART, and coordination of Equipex FIT). NetSys also features an ERC Advanced Grant laureate, two IUF members, and two chairs hosted by UPMC.

The NetSys structure evolved significantly in the recent years. Whisper resulted from a split in REGAL to highlight activities related to the design and improvement of infrastructure software. Almasty emerged as a chair hosted by UPMC on cryptology, to become a full research group gluing cryptology related activities.

Over the past period, six young UPMC or CNRS faculty in the department have been promoted in other Universities, demonstrating the Department ability to boost their carriers. In the meantime, NETSYS recruited four junior UPMC or CNRS faculty, and one senior UPMC faculty.

2. Politique scientifique

The scientific vision of the NETSYS department can be described as the convergence of **Networks** and **Systems**, whose ubiquitous and pervasive aspects can only be mastered when **Security** comes as a foundation for those. Figure 1.1 pictures NETSYS's three edges and maps its constitutives research teams accordingly.

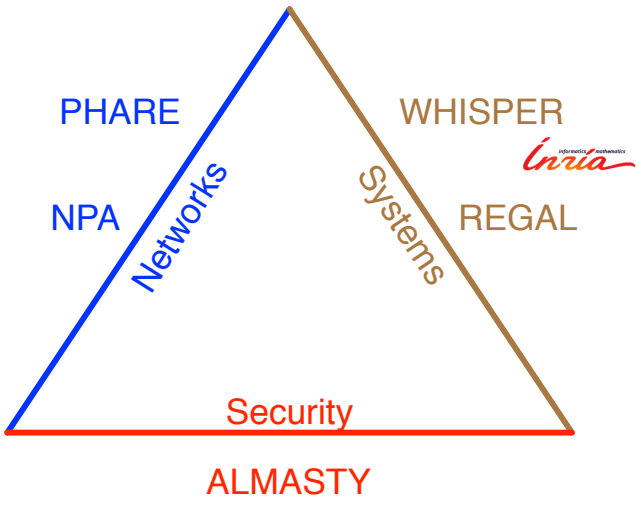


FIGURE 1.1 – The NETSYS Department

Aside from specific research actions that are carried out inside teams and described in the sequel, the NETSYS department developped several joint projects, some of which with visible impact : LIP6 projects (within the department and with other departments, permitting to hire co-advised interns), ANR projects spanning multiple

teams, co-publications spanning teams in all LIP6 departments, joint projects with other UPMC laboratories within the SMART Labex, etc.

The NETSYS department defined clear priorities for temporary and permanent faculty hirings, with a consistent vision over time to promote scientific excellence. Visiting professors frequently collaborate (and later co-author publications) with at least two teams in the department. Each team has its own seminar, open to all members of the department (and there are indeed attendees of other teams when the topic fits). We plan to further strenghen scientific animation by dedicating one or two full days each year to a specific topic that spans multiple teams.

NETSYS research expertise is also a major driving force in several UPMC master programs (KIC EIT Digital ; Master RES – including an international instance in Vietnam, SAR, and SFPN) and in industrial collaborations (ATOS/Renault funded chair on “Smart mobility” ; joint industrial research projects and co-advised PhD students ; creation of several start-up, etc.)

At the national level, besides nation wide research projects, NETSYS participates to LINCS (a joint laboratory with Telecom ParisTech, INRIA, Nokia Bell Labs, and SystemX IRT), has several members in the scientific committee of the CNRS RSD (*Réseaux et Systèmes Distribués*) GDR, and one of the department coordinated EQUIPEX received National Research Infrastructure (IR) label.

International recognition of the department activities reflect on pretigious journal editing committees membership, top-tier conference chairing and organizing, and coordination of large-scale international research projects.

EQUIPE ALMASTY

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The Almasty team was created in April 2015. It was initially formed of a single permanent (A. Joux) with his PhD students. Before the official creation of the team, the persons that initially formed it were not affilliated to any team in LIP6. for the purpose of the scientific assesment, the production of team members during the pre-historic period will thus be assigned to Almasty. The initial scientific theme was mainly based on the advanced ERC grant ALMACRYPT (Algorithmic and Mathematical Cryptology) that was granted at the time of creation and officially started in January 2016.

At the end of 2016, Professor J.C Bajard joined the team together with his PhD students and added the theme of efficient computer arithmetic for cryptographic use. This was, in particular, motivated by an existing scientific collaboration in the H2020 project HEAT on homomorphic encryption.

In September 2017, it is planned that Damien Vergnaud, a newly hired professor, will also join the team. He brings with him his own thematic on the production and use of random and pseudo-random values for cryptographic and security purposes.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	—	—	(1)	1	1	2 (3 à compter 1/9)
Researchers	—	—	(0)	0	0	0
PhD students	—	—	(2)	2	4	6
Others	—	—	(0)	0	0	1
Total	—	—	(3)	3	5	9

Grants (kE)	2012	2013	2014	2015	2016	2017
International	—	—	—	188	738	738
National	—	—	—		63	63
Industry	—	—	—			
Others	—	—	—			
Total	—	—	—	188	801	801

Politique scientifique

The Almasty team is mainly oriented toward the use of algorithmic and effective mathematic techniques to propose and improve modern cryptosystems and assess their security levels. Toward this goal, it makes use of numerous mathematical tools and develop breakthrough algorithms to improve computations well beyond what is reachable with state-of-the-art methods. A dominant part of the mathematical tools used in this context come from number theory and algebraic geometry.

2. Produits de la recherche et activités de recherche

Bilan scientifique

During the past contractual period, the team proposed several important contribution in several direction within its field of research.

Algorithmic number theory.

Computation of discrete logarithms in finite fields. This topic of discrete logarithms has been very active during the past few years. Concerning small characteristic finite field, the use of Frobenius representation together with systematic relation contruction which started in 2013, has been greatly simplified and improved in the article [**ALMA-C9**].

In large characteristic, despite the fact that the computations remain much harder than in the small charateristic case, several asymptotic improvements were proposed by Cécile Pierrot during her PhD, see [**ALMA-J2**] and [**ALMA-C10**].

Computation of class groups of number fields. Computing the class group of number fields is a task which presents a lot of similarity with factoring and the computation of discrete logarithms in prime fields using the number field sieve (NFS) method which as its name indicates relies on the use of such number fields.

However, there is a couple of noticeable differences. First, NFS simultaneously makes use of two (or more) number fields, with often a specialization of one of them as the rational field. Second, the number fields used in NFS are chosen at the start of the algorithm in order to present a nice balance between their degree and size of coefficients. When computing class groups, it is not the case, there is a single, fixed number field, given by a definition polynomial. Depending on its degree and size of coefficients, the cost of computation may vary enormously.

This is the main topic of Alexandre Gélin's PhD, whose defence is planned in September 2017. In a first publication [**ALMA-J5**], an algorithm was proposed to replace the given definition polynomial with an alternate one having smaller coefficients. Whenever this is possible, it greatly speeds up the subsequent class group computation.

Computation of generators for principal ideals in cyclotomic fields. Ideals of cyclotomic fields are used in a number of recent public key cryptosystems. Among the hard problems that underly the security of these systems, we find the problem of recovering a small generator (if it exists) of a given principal ideal.

In truth, this is largely related to the aforementioned computations of class group. The paper[**ALMA-C3**] applies these methods to the case of cyclotomic fields to get a much faster resolution than previously available.

Algorithms for lattice reduction. Lattices and lattice reduction have been widely used for cryptanalytic work since the discovery of the famous LLL algorithm of Lenstra, Lenstra and Lovász in the early 80s. More recently, it appeared that lattice reduction truly becomes a difficult problem when the dimension of lattices grows. This allowed for the introduction of cryptographic schemes based on this hardness.

Despite the high performance of LLL and the numerous improvements that have been discovered since 1982, a lot of open questions remain concerning lattice reduction algorithms. In particular, the issue of the necessary precision for the intermediate computations of orthogonalized bases that are done using approximations for performance reasons is not completely understood.

In an ongoing work [**ALMA-P3**], we propose to use interval arithmetic in this context. The lattice reduction code that arise from this idea is available on Thomas Espitau's webpage.

The team also has a number of publications that aim at improving lattice based cryptosystem through the use of well-chosen representation systems of integers.

Pre-publications. In this topic of algorithmic number theory, the team also has some work in progress, some of which has been announced in the form of preprints. In particular, lets quote an improvement for the choice of elliptic curve in the ECM factoring algorithm [**ALMA-P4**] and a new approach for solving Boolean systems of polynomial equations [**ALMA-P5**].

Security assessment of cryptographic systems.
In cryptography, pinpointing security flaws in the design of cryptosystems rather than in the underlying hard problems is also an important avenue of research. During the past period, the team also published some contributions of this kind [**ALMA-C11, ALMA-C7**].

Linear algebra algorithms.

Linear algebra is an essential ingredient to a large variety of algorithms which are used to solve cryptography-related hard problems. It differs from many other applications of linear algebra by the key fact that, in this case, it needs to perform exactly, i.e., on matrices whose coefficients belongs to finite fields (or sometimes finite rings). As a consequence, approximate methods that converge toward the desired solution are often excluded in this context. However, some exact algorithms get their inspiration from approximate methods, especially where sparse matrices are concerned.

Two recent papers from the team are part of this topic, one deals with the case of nearly-sparse matrices [**ALMA-L3**], the other proposes an efficient matrix-vector verification method [**ALMA-C4**].

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	—	—	3	1	2	2	8
Ouvrages	—	—	2	0	1	1	4
Conferences	—	—	4	1	2	4	11
Logiciels	—	—	0	0	1	0	1
Brevets	—	—	0	0	0	0	0

Sélection des produits et des activités de recherche

Faits marquants

During the last year, two PhD thesis were defended within Almasty, those of Cécile Pierrot and Chrysanthi Mavromati. During the hiring period for research and teaching positions that occured at the beginning of 2017, the first named person was ranked first on two separate positions: CR at INRIA Lorraine and MCF at Limoges' university. The second named person was hiring by the government agency called ANSSI (agence nationale de la sécurité des sysèmes d'information).

3. Analyse SWOT

4. Projet scientifique à cinq ans

The scientific project of the team for the five coming years are based on its research projects, either ongoing or in preparation. More precisely, the relevant projects are:

- ERC Almacrypt. This project is hosted by UPMC, with Institut Fourier as a partner.
- ANR ARRAND Project. This project is hosted by UPMC, with partners from Institut Fourier, Toulon's university and Rennes' university.
- H2020 project Antique (currentlyevaluation phase). This project would be hosted by UPMC with academic and industrial partners from Belgium, France, Germany, Portugal and Switzerland.
- Damien Vergnaud's project (in preparation) on random and pseudo-random generation and its cryptographic uses.

Expanding on the topics already mentioned in the scientific production section, we list below,some of the research direction from the above projects.

Algorithmic number theory.

- Factoring of integers and polynomials over finite fields.
- Discrete logarithms over finite fields.
- Elliptic curves. Discrete logarithms and isogeny search.
- Lattice reduction.
- Polynomial systems resolution over finite fields.

Efficient representation of integers and polynomials.

- Residue number systems for integers and modular integers.
- Residue number systems for polynomials.
- Fast algorithms for residue number system representations.
- Randomized number systems and their use for side-channel protection.

Randomness and pseudo-randomness in cryptography.

- Use and processing of (possibly biased) random sources in cryptography and security.
- Randomized algorithms and derandomization in the cryptographic context.
- Pseudo-random generators, especially those based on hard number theoretic problems.

Security assessment of cryptographic systems.
NIST recently launched a competition in order to standardize new public-key cryptosystems capable of withstanding the potential threat posed by quantum computers. The study of these systems and their security will be part of the core thematics of the team in the coming years.

Proposal of cryptographic systems.
Almasty can also propose some new cryptographic systems and recently wrote two such proposals. One as part of the H2020 project HEAT [**ALMA-C1**]. The other, which is currently a prepublication [**ALMA-P1**], in the context of the Almacrypt ERC project. Improving these systems and assessing their deployability is also an important line of work for the coming years.

EQUIPE NPA

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The Networks and Performance Analysis (NPA) team of LIP6 maintains a tradition of top-notch research in networking since the creation of the laboratory in January 1997. The team's core activities significantly evolved since then – they now go beyond the design of communication protocols to a much wider scope involving highly novel communication paradigms, network governance, secure distributed algorithms, and ultra-large experimental testbeds for the internet of the future. The NPA team is located at the Jussieu campus, aisle 26-00, first floor.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	11	13	13	13	13	13
Researchers	3	3	2	2	2	3
PhD students	21	20	22	19	20	16
Others	11	12	11	12	9	7
Total	46	48	48	46	44	39

Grants	2012	2013	2014	2015	2016	2017
International	633	827	820	1082	876	921
National	586	397	304	244	238	225
Industry	164	287	288	284	256	255
Others	0	0	0	5	5	5
Total	1383	1510	1412	1615	1375	1406

Politique scientifique

NPA team's goal is to propose disruptive solutions and building blocks that are likely to compose the internet of tomorrow. In our vision, future communication systems will be characterized by hyper-connectivity and will require innovative, highly adaptive, and secure approaches. We explore alternative communication models and propose to rely on a thorough understanding of heterogeneous interaction patterns which will be inherent characteristics of future networked systems. The team develops a modern approach of research in networking, through basic research and transfer activities, in strong cooperation with worldwide academic partners and industrial leaders. The core of our work addresses problems related to mobile networked systems, resource modeling and management, scalability, self-stabilization, metrology, and content networking. The team also develops multidisciplinary research in the area of social sciences, governance, and usage.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Research topics. NPA gathers researchers with different skills and specificities towards the ultimate goal of designing, modeling, analyzing, and experimenting building blocks of the future internet. During the reporting

period, the team produced knowledge in several areas of networked systems, both from theoretical and practical points of view. Illustrative examples of research achievements are :

– *Modeling and performance evaluation of networked systems.* The objective is to revisit the way technologies are evaluated and benchmarked by proposing novel performance evaluation tools (based on the latest developments in queuing theory) that are able to tackle the complexity of traffic profiles in future mobile networks. These tools are to be carefully adapted to the different technologies discussed within 3GPP, and then used to benchmark these technologies and perform pertinent choices among them. In the context of software-defined networking (SDN), NPA focuses on network functions virtualization (NFV) performance and dynamicity. We develop innovative models to dimension the low-level resources (e.g., CPU, RAM, I/O) associated with a virtual machine in order to meet performance requirements. Additional expertise and production within the team is the performance evaluation of cellular networks whose positions are modeled by Point Processes. With the use of stochastic geometry, the coverage and throughput of cooperative cellular networks can be evaluated, where the signals can be jointly transmitted through coordinated-multipoint techniques.

– *Radio networks and 5G.* NPA has contributed with significant work under the umbrella of advanced radio networks. In the area of cognitive networking, we proposed a hybrid spectrum sharing strategy that includes novel static and dynamic spectrum sharing algorithms based essentially on a rate compensation approach and adapted best fit algorithms. The proposed hybrid spectrum sharing demonstrates its effectiveness in terms of improving the overall service satisfaction and reducing the number of handoffs while the achieved rate is fulfilling compared to the optimal. We also considered the problem of uplink scheduling under practical constraints imposed by the LTE/LTE-A standard. We proposed a novel scheduling scheme which is standard-compliant and covers a wide range of scheduling objectives. Towards the 5G, the team conducts research on the "slicing" paradigm, which allows logically-isolated network partitions (with a slice being considered as a unit of programmable resources such as network, computation, and storage). One of our objectives is to bring the slicing concept into reality and to develop new models to evaluate potential performance gains.

– *Distributed algorithms.* NPA is very active in designing distributed algorithms for challenged networks, where unexpected topology changes, faults, and attacks of arbitrary nature hit the network. Our focus has been on the spanning network topology and various distributed building blocks when faults occur. To manage that, we proposed self-stabilizing and fault-tolerant algorithms. Self-stabilization deals with the design and analysis of distributed algorithms in which processes are subject to transient failures modifying the content of their variables. We focused on Byzantine faults. This model of faults permits to tolerate the classical transient faults like the corruptions of variables, but also the dynamicity of the network (the change of the topology or the metrics evolve in time). Moreover, this model permits also to manage the faults generated by the attackers. The proposed algorithms handle not only the transient faults but also the quality of the solution, the used memory, and the time of convergence. In this context, we proposed optimal solutions for problems like election leader, mutual exclusion, distributed storage, minimum spanning tree, and minimum degree spanning tree.

– *Swarms of autonomous mobile robots.* We explore solutions to provide a cartography of a damaged area in disaster situations and set up geolocation service using crowdsourcing. In this context, we focus on two main research directions ; the used models and the problems that that industries or military want to solve with swarm robots. There exist a huge variety of robots, each variety implies different capabilities. By capabilities, we mean the vision, the sense of direction, the ability of movement, and the space memory, among others. Each capability implies different models. We studied different models, the goal is to understand what kind of problem we can solve with powerless models, like oblivious robots, robots without chirality, and limited visibility. We proposed solutions for classical problem like gathering, exploration, capture of intruders, different kind of patterns.

– *Connected car.* NPA hosts a professorship chair sponsored by Atos and Renault on the connected car. The outcomes of this chair are manifold, covering topics such as opportunistic networking, named-data networking, and pollution measurement on the move. As an example, the team has proposed, designed, implemented, and experimentally evaluated CarFi (https ://www.carfi.mobi), a patented technology that enables vehicle-to-infrastructure connectivity via Wi-Fi, improving performance while saving on cellular data plans. Thanks to an improved scanning algorithm and numerous optimizations to the connection setup, CarFi makes downloading large amounts of data from a moving car feasible. CarFi is being adopted by companies as a preferential solution to help them reduce costs with cellular operators.

– *Massive data offloading.* Today's networks have to cope with an unprecedented data overload. One of the most promising approaches to accommodate such a data avalanche is to rely on *offloading techniques*. We have

considered two main scenarios : cellular and backbone offloading. In the former case, we propose to alleviate the burden on the cellular network by leveraging unused bandwidth through device-to-device communications. The idea is to finely control the distribution of popular content in urban scenarios by injecting copies using the cellular network only when needed. In the case of backbone offloading, we advocate the use of conventional vehicles equipped with storage devices as data carriers. We proposed an embedding algorithm that computes an offloading overlay where each logical link spans over multiple stretches of road from the underlying road infrastructure. The numerical results show that 20% of vehicles in circulation in France equipped with only one Terabyte of storage can offload Petabyte transfers in a week. Also, caching policies have been developed for the edge-caching problem that make use of multi-coverage events. These policies that can be either periodically proactive or user-triggered and dynamic, offer high performance benefits compared to standard approaches, such as the LRU, or cache the most popular content policies.

– *Software-defined networking.* We have been using the paradigm of software-defined networking in a number of our activities, and in particular on the management of wireless networks. Firstly, self-configuration issues of multi-hop wireless networks are addressed through a collaboration with CEA-List and realization real testbed implementations. Secondly, we show through the analysis of popular and realistic traffic such as adaptive video streaming that centralized management can bring significant performance gains in both wireless home and enterprise networks. Thanks to our studies, we propose resource allocation strategies to avoid instability and unfair resource allocation among competing video players.

– *Content networking.* NPA has developed significant work in this area since more than a decade. One of the topics is P2P-TV, in which we focus on the efficient delivery of huge amounts of video using peer-to-peer techniques. Several approaches were developed to analyze such applications : traffic analysis under controlled usage and user behavior analysis with multiple vantage points. These works rely on passive metrology.

– *Internet governance.* The focus of our research in this area was on identifying and characterizing the roles of the multiple actors, as well their strategies, in the governance of the global internet. In particular, we led a fifteen-year long study that considered not only the main stakeholders but also less explored actors, such as intergovernmental organizations, technical communities and other less organized categories. The outcomes of our study establish the basis of a fine understanding of internet governance at a global scale.

– *Experimental testbeds.* Another major activity concerns the design, development, and operation of large-scale testbeds to accelerate the research of the networking and distributed system community. This initiative started ten years ago with the vision to build an Internet of testbeds related to the Internet technologies and services. Practical outcomes of this initiative include PlanetLab Europe and OneLab-FIT, a facility made up of individual testbeds, which together offer experimentations across heterogeneous resources federated through an access portal. OneLab-FIT is an Equipex that has recently received National Research Infrastructure (IR) label. Overall, the OneLab initiative coordinates (or has coordinated) 10 European/national projects and participates (or has participated) in more than 20 other projects. Another important outcome of our activities on testbeds, we have been using large research facilities for education. In particular, we develop of FIRE adaptors and portable widgets to integrate in different educational systems (e.g., web, e-book, LMS) and use them in two M.Sc. teaching units (“Computer Networks” and “Network Metrology”) as well as in a MOOC.

Collaborations, funding, and international visibility. The team is very active in setting up and conducting collaborative projects, managing an annual budget of around one million euros. Projects are mainly funded by the European Commission, the French National Research Agency (ANR), as well as by industry. Examples of industrial partners of the team are Thales, Microsoft Research, Nokia, Cisco, Atos, Renault, Google, Telefonica, Technicolor, Orange, Toyota, and IBM. The team is also deeply involved in the Excellence Laboratory SMART. The team is highly attractive at the international level ; we maintain continued collaboration with internationally renowned researchers. We regularly host international visitors (students, professors) and send our students as interns in top-level research groups worldwide. Since 2010, the team has the honor to host Paul Mockapetris, internet pioneer and inventor of the DNS (Domain Name System). At the national academic level, the team maintains strong collaborations with leading research groups at both universities and tier-1 research institutions. NPA participated in the setup of the Japanese-French Laboratory for Informatics, a CNRS International Joint Unity between France and Japan. The team was also part of the founding members of the Laboratory of Information, Networking, and Communication Sciences (LINCS). Members of the NPA team hold responsibilities at the highest levels, including UPMC vice-presidency and advisory body of the French Ministry of Research. One of the members of the team was recently a member of Institut Universitaire de France (IUF).

From research to teaching. All the faculty members of the NPA team are involved in the leadership and teaching activities related to the major on Computer Networks at UPMC, which is the most popular in terms of applicants and registered students. This major offers several tracks including dual education for apprentice students and a research oriented track. The NPA team also has a long experience in the international externalization of UPMC education offer. The technical major in Computer Networks has been being duplicated in Vietnam in the framework of the Pôle Universitaire Français since 2006. In Europe, we are the leading university of the technical major in the framework of the European Institute of Innovation & Technology. Furthermore, a member of the team was a co-author of a MOOC on the topic of internet measurements with more than 1,900 people enrolled. This course served as the basis of a full-day tutorial in ACM Sigcomm 2016. Finally, in the context of the FORGE project funded by the European Commission, the team relies on its expertise with experimental testbeds to create richer learning environments to do in a network measurements course.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	13	13	10	15	15	9	75
Ouvrages	4	2	0	3	4	2	15
Conferences	40	45	35	36	45	11	212
Logiciels	5	2	0	1	2	0	10
Brevets	0	0	1	0	0	0	1

Sélection des produits et des activités de recherche

Selected publications :

— A. Ragaleux, S. Baey, and K. Karaca, “Standard-compliant LTE-A uplink scheduling scheme with quality of service”, to appear in *IEEE Transactions on Vehicular Technology*, 2017.
— N. Malouch and A. Abdelfattah, “Modeling and performance analysis of Wi-Fi networks coexisting with LTE-U”, *IEEE Infocom*, Atlanta, GA, USA, May 2017.
— L. Blin and S. Tixeuil, “Compact deterministic self-stabilizing leader election on a ring : the exponential advantage of being talkative”, to appear in *Distributed Computing*, 2017.
— B. Baron, P. Spathis, H. Rivano, and M. Dias de Amorim, “Offloading massive data onto passenger vehicles : Topology simplification and traffic assignment, *IEEE/ACM Transactions on Networking*, vol. 24, n. 6, pp. 3248–3261, December 2016.
— S. Bonomi, A. Del Pozzo, M. Potop-Butucaru, and S. Tixeuil, “Optimal mobile byzantine fault tolerant distributed storage”, *ACM PODC*, Chicago, USA, July 2016.
— T. Begin, B. Baynat, I. Guérin Lassous, and T. Abreu, “Performance analysis of multi-hop flows in IEEE 802.11 networks : A flexible and accurate modeling framework”, *Performance Evaluation*, Vol. 96, pp. 12-32, February 2016.
— J. Auge, T. Parmentelat, N. Turro, S. Avakian, L. Baron, M. A. Larabi, M. Y. Rahman, T. Friedman, and S. Fdida, “Tools to foster a global federation of testbeds”, *Computer Networks*, vol. 63, n. 22, pp. 205–220, April 2014.
— J. Whitbeck, M. Dias de Amorim, V. Conan, and J.-L. Guillaume, “Temporal reachability graphs”, *ACM Mobicom*, Istanbul, Turkey, August 2012.

Instruments/software :

— Experimental platform OneLab (onelab.eu)
— Experimental platform FIT (fit-equipex.fr)
— PlanetLab europe (planet-lab.eu)
— CarFi (carfi.mobi)
— Canarin-II (canarin.net)

Faits marquants

— Publications of results in top-tier conferences and journals – ACM PODC, ACM Mobicom, IEEE Transactions (Vehicular Technology, Networking, Mobile Computing, Parallel and Distributed Systems,

Services Computing, Intelligent Transportation Systems, Networks and Service Management, Network Science and Engineering), IEEE Infocom, Distributed Computing, Computer Networks, Performance Evaluation, Algorithmica, Communications of the ACM.
- Organization of four major international conferences in the areas of mobile and distributed computing : DISC 2016 (general chair and organization), ACM Mobicom 2015 (general chair and organization), ACM Mobisys 2015 (general chair), and ACM PODC 2014 (organization).
- Active participation in administrative and governmental instances – vice-presidency of UPMC (Europe and International Affairs) and scientific advisory at the Ministry of Higher Education and Research.
- Launch of the OneLab experimental platform in 2014.
- Leading position in experimental infrastructures for the future internet.
- Signature of two collaboration agreements between the OneLab consortium and Asian institutes : (i) with the Institute of Computing Technology of the Chinese Academy of Sciences and the Chinese Industry Innovation Center for Future Network and, (ii) with the Taiwanese Institute for Information Industry (iii) (in the presence of Jong-Chin Shen, Deputy Minister of Economic Affairs).
- "Étoiles d'Europe" prize for FP7 OpenLab, 2014. It is worth mentioning that the team also participated as a core member of FP7 MOTO, which also received the same prize in 2016.
- Equipex FIT became a research infrastructure and is now part of the French National Strategy on Research Infrastructures (appeared in the 2016 edition).

3. Analyse SWOT

Strengths. Publications in top-tier journals and conferences. Complementary skills and expertise within the team (from theory to practice). International visibility. Solid capacity to operate and manage international experimental platforms. Production of open software for piloting and exploiting experimental platforms. Solid bridge between research and teaching. Interdisciplinarity (computer and social sciences).

Weaknesses. Difficulty to renew funding for a number of topics we pioneered. Lack of a sufficiently large and stable group of developers. Difficulty to compete (in terms of salary) to hire students and engineers. There is still room for improvement with regard to industrial technology transfer. High administrative overhead.

Opportunities. Recent patents produced by the team open new possibilities for valorization (industrial contracts and start-ups). Links with leading industrial partners has improved in the latest years.Possibility to make FIT/OneLab as the reference European platform for testing new Internet advances. Further explore experimental platforms for education.

Threats. Difficulty to maintain a high-level team of engineers to set-up and maintain experimentation platforms. The time dedicated to experimental platforms, though highly useful, is difficult to value as a research activity. Without financial support, some topics developed by the team risk to disappear.

4. Projet scientifique à cinq ans

5G and beyond. We will propose models for intra- and inter-slice resource sharing in 5G networks. We will investigate variants of Load Based Equipment (LBE) in the context of Listen Before Talk (LBT) sharing methods and introduce more advanced parameters of the physical layer in order to further enhance the performance of future wireless systems. We just started a collaboration with Nokia on this topic. Other topics include the investigation of spectrum aggregation techniques (an important focus of 3GPP) and the interactions between LTE-U/LAA and Wi-Fi when operating within the same unlicensed bands. With this regard, we plan to leverage collaborations with other teams in the laboratory (e.g., operational research).

Performance of wireless networks under spatiotemporal dynamic traffic. This work continues the line of work of the last years on applications of stochastic geometry and queuing on wireless networks. There are several aspects of 5G wireless networks that can appropriately be modeled by these tools. This is because new generation networks will be very dense and heterogeneous. Another important aspect is the requirement of cooperation/coordination between the numerous wireless nodes. Additionally, coexistence between LTE, Wi-Fi, and millions of machines gives rise to several interesting problems that should be treated in a way that incorporates both the spatial and time dimensions. Traffic is not uniform and varies on both space and time, so that a network cannot be treated only on one instant or on some fixed geographic area. These variations of

traffic and their dynamic nature are the true sources of possible under-functioning for the network. Analyzing such models will give rise to appropriate solutions as well.

Distributed algorithms for ubiquituous networks. A key future topic will be to revisit algorithmic models to better fit actual ubiquitous networks such as body area networks, swarms of mobile robots, and spontaneous smartphone-based networks. Those model refinements mandate profound changes to the algorithmic foundations of many communication protocols, as our preliminary findings indicate important shifts in performance when considering finer-grained models.

Societal challenges and IoT. We will dedicate particular attention to the investigation of IoT solutions in the context of challenged networks such as assistance to disabled citizens. The idea is to leverage recent research achievements of the team to make applicative scenarios achieve highest standards in terms of quality of service and reliability. We also intend to provide resilient and secure protocols for wearable networks with a specific target on medical applications. Recent collaborations of the team with Salpétrière Hospital advocate for the timely research in this topic.

Distributed ledges and smart contracts. The challenge is to design specialized secure blockchains with proven QoS guaranties that can be used as alternatives for the classical client-server architectures. The idea is to extend the existing research in distributed systems, networking and internet governance in order to meet the specific requirements of blockchain based applications : churn, high volume of data, high rate of transactions, security, privacy, and resilience.

Analysis and control of data-oriented telecommunication networks. The goal is to make an efficient use of the available data in order to adapt the network operation. This will depend on how flexible the networks are designed as well as on the nature and quality of the measurements/data gathered and their evaluation. We will focus on two main problems : (i) treatment of large volumes of data (including prediction) and (ii) the way the network should be controlled and adapted (e.g., to decide on the time-scale of actions taken). The area requires tools from telecommunications, computer science, and mathematical sciences.

Cache-equipped networks. So far, we have focused on defining appropriate content placement on network nodes, when a request can be treated by multiple sources simultaneously. Several additions and improvements can now be proposed. On the one hand, the cache-equipped load can be taken into account, so that content placement can also target a fair load distribution in the network. Additionally, in the context of big data, and data-oriented telecommunication networking, external information from social networks can be treated and be used as input to decide on the appropriate content placement strategy, when the content is proactively placed on nodes. Finally, questions on hierarchical caching for tree-like networks can be treated, as well as questions related to popularity prediction under different traffic input. Object size can influence caching policies as well.

Creation of a French national capability in Internet cartography. To create this capacity, we will further develop the expertise already existing within the team. We will also deploy infrastructure, using, at the core, the PlanetLab Europe platform that is run by the team. We will also develop international cooperation, such as an emerging collaboration with the CAIDA Internet mapping center at UC San Diego, and a collaboration with RIPE NCC in the Netherlands.

Internet governance as a diplomatic issue. One challenge will be to guarantee that the governance values of the European Union are reflected in the process of governing the global Internet. A major issue will be to ensure that other countries (in particular EU's neighboring countries) adopt laws and rules that integrate the same values. Our research will focus on the impact of disruption phenomena (e.g., "platformization" and "datafication" of social and economic activities, "multistakeholderization" of internet governance, and "fragmentation" of the internet) on political, normative, socio-economic, and geopolitical international relations.

Advanced experimental testbeds. Our efforts to provide advanced, state-of-the-art experimental testbeds will be continued. We will pay particular attention to the usage of the testbeds – how to extrapolate, build, observe, and validate models, algorithms, and methodologies. We plan to provide tools to help experimenters better scale to larger experiences and handle larger amounts of data. One example is about PlanetLab Europe. Its evolution includes : (i) continued integration into pan-European federation of experimental platforms, including federation with platforms overseas and integration with the GÉANT Testbeds Service. Finally, we will develop the pedagogical potential of experimental testbeds for use in our classrooms and those of others who teach computer networking. We base this on our work having already included PlanetLab Europe-based lab exercises in a course for the Master's program and a MOOC.

EQUIPE PHARE

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The PHARE team was created in 2004 and is led by Guy Pujolle. The team addresses the conception of new network protocols and algorithms to tackle rising challenges in communication networks, going from their theoretical modeling, to their prototyping and experimentation by their implementation or integration in real systems. The focus of the team in the last few years concentrates in the evolution of communication technologies with network virtualization and softwarization technologies.

Tableau des effectifs et moyens de l'équipe

Typologie	2012	2013	2014	2015	2016	2017
Enseignants/chercheurs	5	5	5	5	5	4
Postdocs	4	5	10	8	6	6
Doctorants	25	25	26	22	20	17
Ingénieurs	1	4	5	3	3	2
Total	35	39	46	38	34	29

Grants	2012	2013	2014	2015	2016	2017
International	196.4	202.1	173.3	99.1	53.1	0
National	178.2	270.8	427	665.1	780.4	812.5
Industry	101.2	173.8	197.4	83.4	47.5	31.9
Total	399.2	570.3	721.3	822.5	880.9	844.5

The team is roughly composed by 2/3 of Ph.D. students, besides postdoc fellows and engineers. In addition to these numbers, the team recruits about 15 master interns, and hosts on average two invited researchers per month. It is worth pointing out the departure of an associate professor, Rami Langar, on a full professor position at UPEC in 2016.

The global contract volume on the reference period is 4,239 MEuro, with a non-negligible part coming from industrial contracts (15%) and European projects (17%). National projects are funded by ANR (15%), the Investissement d'Avenir programme (31%), and the remaining part derives from FUI projects.

Politique scientifique

The scientific approach of the team is to go from the conception of new network protocols and algorithms to tackle rising challenges in communication networks, to their prototyping and experimentation by their implementation or integration in real systems. In order to do so, the adopted methodology concerns in algorithm design ranges using game theory and combinatorial optimization in resource allocation problems, data analytics and graph theory in routing and network management problems. Prototyping is often accompanied by an open source project (as for the activities related to LISP, MPTCP, ONOS, for instance) and is often run in collaboration with telecommunication operators (recently with Thales, Orange, Telecom Italia and Ericsson). We can note that, in certain cases, if the prototype shows a real technology advance, the possibility to create a start-up is investigated, in the frame of the SATT, a technology transfer company associated to the university. The last created start-up during the reporting period is Blue Communications.

This scientific approach allows the PHARE team to be distinguished at both national and international levels in its research area. This is proven by the involvement of the team in many national, European and extra-European collaboration frameworks and projects, major conferences, the responsibility in national and international societies (e.g., steering board of the CNRS GdR Réseaux et Systèmes Distribués, chairing of the Internet Technical Committee of the IEEE Communications Society) and in open source project management (e.g., chairing of one brigade of the Open Network Operating System, ONOS).

2. Produits de la recherche et activités de recherche

Bilan scientifique

The main goal of the PHARE team is to experiment network protocols and control algorithms for future networks. Currently, the team concentrates its efforts on the conception of a holistic architecture introducing the necessary functionalities to be operated in future networks.

More precisely, the envisioned architecture includes three levels : the 'Cloud' level with central data-centers, the 'Fog' (vendor-oriented term) or MEC (Mobile Edge Computing) level (provider-oriented term) with data-centers on the network boundaries, and the 'Skin' level with so-called femto-datacenters or wall-datacenters, within a few meters from the user terminal or the connected object. On this three-level network structure, the PHARE team works on a variate set of network communications issues arising when virtualization takes place, for both network and system components. In particular, the involved communication scenariii cover, besides users or objects, virtual machines, which raise novel challenges in terms of resource allocation, the embedding of their virtual networks with the underlaying physical layers, their chaining, their orchestration and their security.

This reference architecture includes several special cases depending on the virtualization scope, the number of layers and their interlacing. For instance, in Cloud-based Radio Access Networks (C-RAN) two levels are used and virtualized components are those network functions related to the radio signal processing. In MEC architectures instead, the virtualized components are the end-servers and possibly load-balancing and routing functions, on a three-level architecture. Completely distributed and autonomic architectures are also envisioned, where the number of levels can dynamically change and where there can be no hierarchy among communicating components at different levels, and where the communicating nodes can be objects such as drones, industry 4.0 sensors or actuators, etc. This multi-level structure leads to a large set of algorithmic and protocol design challenges, which are the core of the PHARE team research activities. These challenges go from theoretical analysis of network protocol features to their experimentation via prototypes to assess their efficiency. In the following part, we describe our vision summarizing different results and grouping them by main interlaced directions.

The first research direction concerns those protocols designed to manage the virtual machine mobility and their virtual network overlay communications, possibly taking into consideration user mobility as well. The team proposed various improvements to virtual network overlay protocols such as TRILL (Transparent Interconnection of a Lot of Links), LISP (Locator/Identifier Separation Protocol) and MP-TCP (Multi-Path Transmission Control Protocol). These protocols have attracted our interest due to their features – based on multipath transmission and multi-level address decoupling and mapping (e.g., within the data-link, network or transport layers) – match the requirements of virtual network communications where path diversity can strongly vary during a communication scenario and where objects, users and virtual machines can be mobile. Moreover, these overlay protocols allow incremental deployment over existing networks, which is an important aspect for the sake of experimentation and reproducibility. In this area, the team works toward the definition of protocol extensions, including also standardization, open source development and integration into programmable network software, and their evaluation in realistic scenarios.

The second research direction is about the virtual machine urbanization, i.e., the process of piloting the orchestration of virtual machines as a function of the varying state of the network segment between them and the user. Once created, virtual machines are chained and eventually urbanized, determining the best possible placement in a multi-level network structure. Once placed, network and computing resources are to be allocated and scheduled while adapting to the operational context. The major difficulty in the definition of corresponding algorithms is the large number of parameters that shall be taken into account and that range from performance metrics to energy consumption levels, from user mobility to network reliability or security targets, etc. A large

number of results have been obtained in the last few years as documented in the publication record. The adopted methodology is often related to combinatorial optimization, game theory, graph theory, data analytics, and reliability theory.

The third axis is about objects communications. Important results were obtained on the mobility of objects equipped with RFID radio components. The team developed rapid RFID localization algorithms for high-mobility scenarios such as industrial chains of production or transportation. The developed algorithms guarantee very high precision under strict temporal constraints. Besides classical publications, this activity permitted to write a patent. This research activity is led in strong collaboration with the Paris-Saclay University (LRI lab), and a start-up of the team, Green Communication, which focuses on participatory networking. Very recently, another start-up, Blue Communications, was created by the team with the goal of developing a network of drones with adapted communication technologies, including scenarios where the drones embed femto data-centers and resources, forming a flying cloud. The flying cloud is realized by a set of connected drones capable of positioning themselves to accomplish a given mission and capable to taking charge of every type of tactical or entertainment mission thanks to the computing power of the embedded femto datacenters.

The fourth axis is about network reliability and security. Important results were obtained on the reliability analysis of data-center architectures, with proposals of algorithms to support a high geographical distribution of data-center facilities. Another recent research activity is on the computation of availability levels for virtualized network services, in order to instruct orchestration algorithms for the corresponding virtual machine instances, including novel protection schemes proposed by the team. Our attention was recently also devoted to securing the cloud orchestration interface between cloud network stack controllers and virtualization servers ; a solution based on blockchain approach was conceived and lead to a patent proposal. Finally, in this area the team has strong cooperation with the network security team of Telecom ParisTech on the usage of secure elements for securing the communications of both terminal machines and virtual machines, by embedding the secure elements in user devices and virtualization system elements.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	13	4	12	6	9	5	49
Ouvrages	0	1	1	2	1	2	7
Conferences	21	18	23	11	18	8	99
Logiciels	1	0	1	0	2	2	6
Brevets	0	1	1	2	2	1	7

These numbers reflect what already exposed and the research production shows an interesting range with books, journals, conferences from the one hand, and software and patents from the other hand.

Sélection des produits et des activités de recherche

— A. Achour, K. Haddadou, B. Kervella, G. Pujolle, "A SIP-SHIM6-Based Solution Providing Interdomain Service Continuity in IMS-based Networks", *IEEE Communications Magazine*, vol. 50 (7), pp. 109-119, 2012.

— S. Hoteit, S. Secci, R. Langar, G. Pujolle, "A Nucleolus-based Approach for Resource Allocation in OFDMA Wireless Mesh Networks", *IEEE Transactions on Mobile Computing*, vol. 12 (11), 2013.

— S. Secci, G. Pujolle, TMT. Nguyen, S. Nguyen : 'Performance-Cost Trade-off Strategic Evaluation of Multipath TCP Communications', *IEEE Transactions on Network and Service Management*, vol. 11 (2), pp. 250-263, 2014.

— R. Langar, S. Secci, R. Boutaba, G. Pujolle, "An Operations Research Game Approach for Resource and Power Allocation in Cooperative Femtocell Networks", *IEEE Transactions on Mobile Computing*, vol. 14 (4), pp. 675-687, 2015.

— A. Amokrane, R. Langar, R. Boutaba, G. Pujolle : 'Flow-Based Management For Energy Efficient Campus Networks', *IEEE Transactions on Network and Service Management*, vol. 12 (4), pp. 565-579, Nov. 2015.

— S. Secci, P. Raad, P. Gallard : 'Linking Virtual Machine Mobility to User Mobility', *IEEE Transactions on Network and Service Management*, vol. 13 (4), pp. 927-940, 2016.

— V. Del Piccolo, Ah. Amamou, K. Haddadou, G. Pujolle : 'A Survey of Network Isolation Solutions for Multi-Tenant Data Centers', *IEEE Communications Surveys and Tutorials*, vol. 18 (4), pp. 2787-2821, Dec. 2016.
— A. Ceselli, M. Premoli, S. Secci : 'Mobile Edge Cloud Network Design Optimization', *IEEE/ACM Transactions on Networking*, Vol. 25, No. 3, pp : 1818-1831, June 2017.

Faits marquants

— Relevant number of publications in top journals in the communication networks area (IEEE/ACM Transactions on Networking, IEEE Transactions on Mobile Computing, IEEE Transactions on Network and Service Management, IEEE Network, IEEE Communications Magazine) as well as seven books ;
— G. Pujolle nominated to the Institut Universitaire de France (IUF) ;
— G. Pujolle awarded the grand prix Orange 2013 of the French Academy of Science ;
— Strong technology transfer activity with six filed patents during the reporting period, some of which already accepted as international patents ;
— S. Secci acts as chair of the Internet Technical Committee (ITC), joint committee between the IEEE Communications Society and the Internet Society (ISOC) ;
— Coordination of many national projects (ABCD, LISP-LAB) and European projects (SECFUNET, SGS, SOFNETS) ;
— Participation to the steering board of the CNRS GdR Réseaux et Systèmes Distribués (RSD) and co-chair of its transversal action on virtualization (S. Secci).

3. Analyse SWOT

The strong point of the team is the homogeneity of its members and the ability to cover a large R&D width, that is, to conceive novel architectures, protocols, control algorithms, and to evaluate them, develop and experiment them and to create start-ups if the conditions for an economical success are met. The team exploits theoretical tools for the network modelling and performance evaluation, develops and tests its proposals and attempts to push it into the economical ecosystem. The PHARE team completes these actions with an important set of publications in the best international journals in the area, with a big number of research contracts, a strong participation to decision-making instances, conference direction, strong collaboration with international research centers and the dissemination of the state of the arts in recognized books.

A point that needs to be enhanced is the lack of visibility on the lifetime of the team that decreases in size because of a departure, toward an external full professorship, of one of its permanent members in 2016, and the next retirement of the team head in 2019. Moreover, administrative constraints are inexorably increasing, in particularly denying the hiring of highly qualified software engineers and limiting the number of Ph.D. students that can be registered, so strongly limiting the international collaboration of the team through Ph.D. co-tutelle.

The possibilities offered to the PHARE team are significant because of the importance of the addressed technological challenges tackled by the team in the economic development, which are not expected to decrease because of the increasingly pervasiveness of network communications in the digital society. The recognition of the team by the community makes it a privileged partner in collaborative research sollicitation from different channels.

The risks associated to this context derive from the increasing administrative overhead that creates problems unplanned at the contract negotiation phase, for high-qualified personnel recruitment and the impossibility to register new Ph.D. students forming the base of research activities. Moreover, the small size of the team generates additional overhead to members in case of external promotion. The impossibility to plan for a new permanent staff recruitment to maintain the size of the team is also a risk factor undermining the team research activities.

4. Projet scientifique à cinq ans

In the forthcoming five years, the PHARE team will continue its research and development activity toward the design of next generation networks. The studies of the last five years were mainly about issues related to the convergence between mobile and fix networks facilitated by network virtualization and the design of programmable network control planes. In future studies for the next five years, the team will henceforth concentrate on

the next generation network challenges, as anticipated by some prospective articles of the team. In particular, network slicing, D2D (Device-to-Device) communications augmentation, coupled with network data analytics, bring novel challenges in the team research agenda. Let us describe these challenges in more details in the following.

First of all, network slicing – i.e., a form to isolate application-driven virtualized access networks encompassing both wireless signal processing functions and standard network functions – requires novel ways to qualify network efficiency in a context where physical resource sharing and resource allocation are expected to be run differently than in legacy systems, while having to guarantee full independency and isolation among slices. For example, a single network system may need to be able to support one slice that could be a large number of objects connected, each with a very low bandwidth. Another slice could be dedicated to a bandwidth-harvesting video-surveillance system requiring coding and cyphering functions. Another slice could be dedicated to Voice over IP calls traffic, etc. Two distinct decision-making dimensions appear in network slicing : intra-slice resource allocation and inter-slice resource sharing. In terms of intra-slice resource allocation, resources of different nature, but correlated with each other, are expected to be allocated jointly in a same orchestration decision, rather than following independent resource allocation processes. There is indeed a dependence on the amount of link bandwidth needed, the computing power required, and the spectrum width, for instance, for a given slice. How to measure fairness in such a multi-resource context is not clear, and the state of the art does give clear directions only for single-resource fairness and allocation rules. In terms of inter-slice co-existence, resource sharing is an important challenge because two slices may have to share the spectrum, the virtualization system resource and the network link resource, for instance. In the radio segment, available frequencies are in the 0-120 GHz band, and these resources are very inefficiently used and shared among users ; there is the need to quickly detect idle frequencies across the different slices so that traffic can be dynamically stitched to those idle, possible tiny and spectrally spread, bands, while understanding the traffic patterns of every slice.

The massive utilization of D2D (Device to Device) communication is expected to offload the infrastructure from bandwidth exposition and energy consumption peaks. With direct communications among users, the gains can be huge because the signals do not pass through intermediate nodes and network/computing resources that can be far and expensive to use (e.g., in terms of energy). This direction leads to the novel paradigm of collaborative networks, that is, networks where the users' node resources are shared to support computing services without depending on far datacenters. In forthcoming 5G networks specified by the 3GPP standardization body, D2D capabilities will be integrated and discovered as 'ProSe (Proximity Services)'. This is already happening in local-area wireless networks with Wi-Fi Direct, for instance. We will in particular address interference management in D2D communications, and its relationship with routing, mobility and IoT management in dense environments.

Finally, the research directions we described will leverage on the existence of a programmable virtualized infrastructure and enhanced control-plane intelligence they support. Such a network artificial intelligence will make use of network data analytics and machine learning algorithms to classify and clusters traffic flows, users, servers, networks and applications, and exploit this knowledge to detect anomalies, anticipate impairment hence programming the network accordingly, going through various network structure levels and acting at the slice network layer as well. These algorithms will optimize sliced network resource allocation in a more efficient way. These algorithms will identify the best nodes to code packets through network coding. They will also support cognitive radio operations in the detection of exploitable frequencies and the optimization of transmission powers of devices and network access points. The final goal is to completely master the networks by an extensive use of self-control, self-configuration, self-recovery, self-securing algorithms. Our goal is to design the algorithms for such networks using artificial intelligence tools and to exploiting them via advanced network operating and orchestrating systems.

Long term researches will address what we refer as 'morphware' (from greek workd 'morph', i.e., self-transforming) networks, a concept that goes beyond so-called cognitive/intelligent networks in that an intelligent network control-plane automatically modifies and adapts the network resources as a function of network demands and changing network and system states. This direction may eventually require some hardware support to virtualized network operations. For example, the operation of virtualized machines and network functions directly at the programmable card chip and programmable processor levels may reveal as a viable compromise between agility due to network softwarization and security and performance granted by the usage of hardware

elements.

EQUIPE REGAL

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

Regal is a joint research team between LIP6 and Inria Paris created in 2005. The team is localized at LIP6. In 2014, the new Whisper team was split off from Regal, focusing on tools for building and maintaining infrastructure (system) software. The results of the members of the Whisper team (former members of Regal) in 2012 and 2013 are not presented in this document and are detailed in the Whisper evaluation document.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	11	10	11	8	7	7
Researchers	4	4	4	2	2	2
PhD students	23	23	14	18	12	13
Others	8	6	6	3	2	2
Total	46	43	32	30	23	24

In 2014, the new Whisper team was split off from Regal. Additionally, two associated Professor (maître de conférences) left the team in 2014 and 2016 to obtain a full professor position at Telecom SudParis and Savoie Univ. An Associated Professor obtained an associated professor position at New-York University Shanghai and left the team in 2015. Two Associated Professors joined the team in October 2012 and October 2016 respectively.

Grants	2012	2013	2014	2015	2016	2017
International	0 K€	719 K€	0 K€	0 K€	0 K€	280 K€
National	505 K€	0 K€	0 K€	0 K€	512 K€	0 K€
Industry	30 K€	0 K€	129 K€	45 K€	0 K€	0 K€
Others						
Total	535 K€	719 K€	129 K€	45 K€	512 K€	280 K€

Politique scientifique

The research of Regal addresses both theoretical and practical issues of *Computer Systems*, i.e., its goal is a dual expertise in theoretical and experimental research. Our approach is a "virtuous cycle" of algorithm design triggered by issues with real systems, which we prove correct and evaluate theoretically, and then eventually implement and test experimentally.

Regal's major challenges comprise communication, sharing of information, and correct execution in large-scale and/or highly dynamic computer systems. While Regal's historically focused in static distributed systems, since some years ago we have covered a larger spectrum of distributed computer systems : multicore computers, clusters, mobile networks, peer-to-peer systems, cloud computing systems, and other communicating entities such as swarms of robots. This holistic approach allows the handling of related problems at different levels. Among such problems we can highlight communication between cores, consensus, fault detection, scalability, search and diffusion of information, allocation resource, replication and consistency of shared data, dynamic content distribution, and multi-core concurrent algorithms.

In in particular, Regal focuses on three key challenges :
— the adaptation of algorithms to the new dynamics of distributed systems ;
— data management on extreme large configurations ;
— the adaptation of execution support to new multi-core architectures.

We should emphasize that these challenges are complementary : the two first challenges aim at building new distributed algorithms and strategies for large and dynamic distributed configurations whereas the last one focusses on the scalability of internal OS mechanisms.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Research Axis 1 : Distributed Algorithms for Dynamic Networks and Fault Tolerance Nowadays, distributed systems are more and more heterogeneous and versatile. Computing units can join, leave or move inside a global infrastructure. These features require the implementation of *dynamic* systems, that is to say they can cope autonomously with changes in their structure in terms of physical facilities and software. It therefore becomes necessary to define, develop, and validate distributed algorithms able to managed such dynamic and large scale systems, for instance mobile *ad hoc* networks, (mobile) sensor networks, P2P systems, Cloud environments, robot networks, to quote only a few.

Axis 1 addresses both system dynamic and fault tolerance through various aspects : (1) modeling, (2) Fault Detection, (3) Self-Stabilization,and (4) Dynamic System Design. Our approach covers the whole spectrum from theory to experimentation. We design algorithms, prove them correct, implement them, and evaluate them in simulation, using OMNeT++ or PeerSim, on large-scale real platforms such as Grid'5000, or using real traces.

Fault Detection. In an ideal world, distributed systems should provide reliable and continuous services despite the failures of some of their components. Obviously, the real world is not ideal and fault may occur. A classical way for a distributed system to tolerate failures is to detect them and then recover. It is now well recognized that the dominant factor in system unavailability lies in the failure detection phase. Regal has worked for many years on practical and theoretical aspects of failure detections. Since 2013, we address both theoretical and practical aspects of failure detector.

In theoretical aspect, we focus on adaptation of failure detector to dynamic environments whose membership is unknown and the path between nodes are built over the time. We studied Ω, the eventual leader election failure detector. This failure detector is known to be weakest failure detector to solve agreement protocols such as Paxos. Nonetheless, as far as we are aware of, no one of these protocols tolerates a free pattern of node mobility. We propose a new protocol for this scenario of dynamic unknown networks. We also show that Ω is the weakest failure detector to implement an eventually consistent replicated service. More recently we study the k-set agreement problem, a generalization of the consensus problem where processes can decide up to k different values. Exploiting the formalism of the Time Varying Graph model, we propose new quorum-based failure detectors for solving k-set agreement in dynamic networks with asynchronous communications.

On Practical aspect, we propose several algorithms to implement efficient failure detection services. We introduce the Two Windows Failure Detector (2WFD), an algorithm that provides QoS and is able to react to sudden changes in network conditions. We also proposed new failure detectors that expresses the confidence with regard to the system as a whole.

Self-Stabilization. A *snap-stabilizing* protocol, regardless of the initial configuration of the system, guarantees that it always behaves according to its specification. As part of the ANR project *SPADES*, we considered the message forwarding problem that consists in managing the network resources that are used to forward messages. We proposed a snap-stabilizing algorithm for the message forwarding problem in tree networks that uses a constant number of buffers per link.

We have also worked on the problem of committee coordination. Moreover, we proposed a silent self-stabilizing leader election algorithm for bidirectional connected identified networks of arbitrary topology. We introduced the notion of *gradually stabilizing* algorithm as any self-stabilizing algorithm with the following additional feature : if at most τ dynamic steps occur starting from a legitimate configuration, it first quickly

recovers to a configuration from which a minimum quality of service is satisfied and then gradually converges to stronger and stronger safety guarantees until reaching a legitimate configuration again. We illustrate this new property by proposing a gradually stabilizing unison algorithm.

Recently, as part of the ANR project *SPADES*, we also studied a kind of prefix tree structure, *a.k.a.*, tries, as a distributed data structure for indexing and retrieving information. As part of the ANR project *SPADES*, we proposed a self-stabilizing maintenance algorithm for a prefix tree using a message passing model.

Research Axis 2 : Distributed Data Sharing information is one of the major reasons for the use of large-scale distributed computer systems. Replicating data at multiple locations ensures that the information persists despite the occurrence of faults, and improves application performance by bringing data close to its point of use, enabling parallel reads, and balancing load. This raises numerous issues : where to store or replicate the data, in order to ensure that it is available quickly and remains persistent despite failures and disconnections ; how to ensure consistency between replicas ; when and how to move data to computation, or computation to data, in order to improve response time while minimizing storage or energy usage ; etc. The Regal group works on several key issues related to replication : replica placement, scalable strong consistency for replicated databases, and theory and practice of eventual consistency.

Regal is studying replication for two main purposes : reliability and performance.

— We have continued our research on data storage reliability, going beyond the context of peer-to-peer distributed hash tables (DHTs). We have studied the impact of data placement on its durability.
— We have proposed algorithms and mechanisms to automatically adapt the replication degree to face the demand. This provides the ability to offer good access performance while limiting resources waste. Our mechanisms also place the copies close to the users to enhance the performance. In a cloud context, we have designed a new geo-replicated key-value store service that automatically reconfigures the set of replicas.
— We have also proposed new distributed structures to efficiently insert and query data in a dynamic environment.

As just mentioned, Regal's contributions cover the whole spectrum of consistency, in both systems- and semantics-oriented conferences.

— We have designed protocols and implementations that provide strong total-order transactional guarantees yet are highly parallel internally.
— We have designed and implemented highly-efficient and scalable *transactional causally-consistent* (TCC) protocols.
— We continued our research on CRDTs, data types designed to provably converge to a meaningful value despite concurrent updates at different replicas.
— Our current focus, dubbed "Just-Right Consistency, " consists of tailoring consistency to the specific requirements of an application, in order to combine the strongest possible consistency guarantees with the weakest possible synchronisation protocol.
— Finally, a recent research direction is understanding and explaining the structure of the consistency design space.

Research Axis 3 : Operating systems Recent years have seen the increasingly widespread use of multicore architectures and virtualized environments. This development has an impact on all parts of the system software and has led us to propose new ways of thinking about Machine Runtine Environments (MREs, a.k.a. language-level virtual machines), system mechanisms, and techniques to make kernel code more scalable and robust. At the user level, we focus on the management of resources, isolation, and concurrency for MREs, supporting languages such as Java. At the kernel level, we focus on improving the memory and cache management in various virtualized environments such as Xen hypervisor or linux-containers targeting big data applications on multicore architectures.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	7	9	4	9	12	1	42
Ouvrages		1	1				2
Conferences	31	39	24	34	21	11	160
Logiciels	0	0	0	0	0	0	0
Brevets	0	0	0	0	0	0	0

Sélection des produits et des activités de recherche

Faits marquants

— Nicolas Geoffray received in 2013 the 2nd prize for the best PhD thesis in Operating System, from the French Chapter of ACM SIGOPS for his thesis titled “Fostering Systems Research with Managed Runtimes”.
— Regal is the leader of the European project SyncFree, started in October 2013
— *Lower bounds and optimality for CRDTs*. This is the first paper to study the inherent lower bounds of replicated data types. The contribution includes derivation of lower bounds for several data types, improvement of some implementations, and proved optimality of others. This result was presented at POPL 2014.
— *Garbage collection for big data on large-memory NUMA machines*. We developed NumaGiC, a high-throughput garbage collector for big-data algorithms running on large-memory NUMA machines. This resul is published at ASPLOS 2015.
— *Explicit consistency*. We propose an alternative approach to the strong-vs.-weak consistency conundrum, *explicit consistency*. This result has been presented at EuroSys 2015. We have also developed a new sound logic for proving the correctness of a distributed database under concurrent updates. This result is published at POPL 2016.
— *The weakest failure detector of implement eventual consistency*. We found the weakest failure detector to implement an eventually consistent replicated service. This theoretical result has been presented at PODC 2015.
— We initiate a collaboration with ICL Lab (University of Tennessee) to study failure detection in Exascale computing. We designed and evaluated a new robust failure detector. This result is published at SC 2016.

3. Analyse SWOT

Force. Our work in distributed systems and algorithms area is well recognized by the international community. We have been invited to be a part of the organizing committee of known conferences in distributed algorithm and dependable communities : PODC, DISC, SSS, EDCC, OPODIS, SBAC-PAD. We have also regularly been invited to serve on the program committees of major conferences in the field such as ASPLOS, DISC, IPDPS, Europar, NCA, ICPP, EDCC, ISSRE. Furthermore, we have publications in some of the best venues (ASPLOS, OSDI, PODC, POPL, EuroSys, Middleware Supercomputing, IPDPS, ICDCS, SRDS ...) and journals (Algorithmica, JPDC, TCS, TPDS ...) in our research area, as well as 25 invited talks.

Our work continues to interest industry, as evidenced by our cooperation with Scality, Magency and with industry partners in the SyncFree and LightKone european projects. Notably, our Conflict-free replicated data types (CRDTs) have been implemented in the cloud database Riak.

Weakness. The competition is strong. We will continue to aim for the best conferences, and need to target journals more aggressively. Given the departure of 3 associated professors, it is not clear how Regal will redefine its research topics after Marc Shapiro’s mandatory retirement, scheduled for April 2019.

4. Projet scientifique à cinq ans

Versatility and heterogeneity are key features of new generation of distributed systems. With the emergence of IoT, edge devices, and cloud computing, there is a radical change of our vision of distributed systems. Fog

computing is a typical example of such hybrid systems. It extends the Cloud paradigm to the edge of the network, thus enabling new services. These systems are characterized by a) Location awareness ; b) Widespread geographical distribution ; c) Mobility ; d) Very large number of nodes ; e) Heterogeneity. To address the above challenges, we propose the following research directions.

Resource management in hybrid systems.

High heterogeneity in terms of CPU power, Storage capacity, networks bandwidth, and reliability is a major change for systems designer. Managing resources on large hybrid systems requires some fully decentralized solutions and to rethink the way various platforms can collaborate with each other. We will focus on data management in such systems.

Algorithm for dynamic systems.

In the area of *Algorithms for Dynamic Environments*. In dynamic environments, the high volatility of nodes makes it impossible to guarantee a consistent view of the system. Decisions must be based on a partial and possibly false view of the distributed system. Algorithms must tolerate this uncertainty. We will tackle this issue using the two following promising approaches.

Indulgent algorithms naturally support uncertainty, relaxing liveness properties. For instance, indulgent consensus algorithms never violate safety and guarantee liveness once the system becomes synchronous. However, these algorithms may never terminate if the system never becomes synchronous. We will study how the use of confidence thresholds can ensure probabilistic liveness properties considering both crash and byzantine behaviours.

Approaches like *self-stabilization* are not satisfactory because self-stabilizing algorithms do not mask the effects of faults during the finite period of time after perturbations (transient faults, topological change, *etc.*) ceases and before the system resumes its correct behavior. In other words, safety is not guaranteed all the time. Quite few variants of self-stabilization have been proposed to offer more stringent guarantees during the convergence phase after the faults cease (*e.g.*, safe-convergence, fault-containment, superstabilization, snap-stabilization, robust stabilization, etc.) However, except a few work (mainly superstabizing algorithms), those variants have not been addressed within dynamic environments.Our second objective thus consists in providing extra safety properties dealing with topological changes over time, using and adapting the variants of self-stabilization. This is quite a challenging objective, even for simple tasks.

EQUIPE WHISPER

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

Whisper was created as a spin-off of the Regal project-team in May 2014. Since Bertil Folliot, Julia Lawall and Gilles Muller are former members of the Regal project-team, this report presents the work they conducted in Regal in 2012-2014 and later in Whisper.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	1	1	1	1	1	1
Researchers	2	2	3	3	3	3
PhD students	5	5	7	3	3	2
Others	1	0	1	2	2	2
Total	8	7	11	9	9	8

Grants	2012	2013	2014	2015	2016	2017
International	1	1				
National	1		1			1
Industry			3		3	
Others					1	
Total	2	1	4		4	1

Politique scientifique

The focus of Whisper is on how to develop (new) and improve (existing) infrastructure software. Infrastructure software (also called systems software) is the software that underlies all computing. Such software allows applications to access resources and provides essential services such as memory management, synchronization and inter-process interactions. Starting bottom-up from the hardware, examples include virtual machine hypervisors, operating systems, managed runtime environments, and standard libraries. For such software, efficiency and correctness are fundamental. Any overhead will impact the performance of all supported applications. Any failure will prevent the supported applications from running correctly. Formal methods are increasingly being applied to operating systems code in the research community. Still, such efforts require a huge amount of manpower and a high degree of expertise, which makes this work difficult to replicate in standard infrastructure-software development.

In terms of methodology, Whisper is at the interface of the domains of operating systems, software engineering and programming languages. Our approach is to combine the study of problems in the development of real-world infrastructure software with concepts in programming language design and implementation, *e.g.*, of domain-specific languages, and knowledge of low-level system behavior. A focus of our work is on providing support for legacy code, while taking the needs and competences of ordinary system developers into account.

We aim at providing solutions that can be easily learned and adopted by system developers in the short term. Such solutions can be tools, such as Coccinelle for transforming C programs, or domain-specific languages such as Devil and Bossa for designing device drivers and kernel schedulers. Due to the small size of the team, Whisper mainly targets operating system kernels with a focus on multicore architectures, and runtimes

for programming languages. We put an emphasis on achieving measurable improvements in performance and safety in practice, and on feeding these improvements back to the infrastructure software developer community.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Research axis 1: Tools for improving legacy infrastructure software A cornerstone of our work on legacy infrastructure software is the Coccinelle program matching and transformation tool for C code. Coccinelle has been in continuous development since 2005. Today, Coccinelle is extensively used in the context of Linux kernel development, as well as in the development of other software, such as wine, python, kvm, and systemd. Coccinelle differs from other matching and transformation tools such as CIL for C code and Spoon, developed in the Inria Spirals team for Java code, by its entire reliance on concrete syntax, which is familiar to developers, its ability to reason about relationships between scattered code fragments and its ability to perform transformations on them. Coccinelle differs from program analysis tools such as Frama-C by sacrificing soundness and completeness, thus reducing running times and improving ease-of-use, while at the same time making it easy for the user to inject his knowledge about the specific properties of the target software into the rule base.

Currently, Coccinelle is a mature software project, and no research is being conducted on Coccinelle itself. Instead, we leverage Coccinelle in other research projects, both for code exploration, to better understand at a large scale problems in Linux development, and as an essential component in tools that require program matching and transformation. The continuing development and use of Coccinelle is also a source of visibility in the Linux kernel developer community. We submitted the first patches to the Linux kernel based on Coccinelle in 2007. Since then, over 4500 patches have been accepted into the Linux kernel based on the use of Coccinelle, including around 3000 by over 500 developers from outside our research group. Additionally, Julia Lawall has been invited to present at the Linux Kernel Summit, the KVM Forum, the Linuxcon plenary kernel developer panel, and the Linux Security Summit based on her work on Coccinelle, and we have hosted kernel developers Luis Rodriguez (Suse) and Greg Kroah Hartman (Linux Foundation) for longer term visits. Coccinelle has also benefited from engineering support from Inria, making it possible to improve the overall user experience (language features, build system, etc.).

Our initial work in the evaluation period centered around the wrapping up of several PhD theses. In 2013, Suman Saha completed a PhD on Hector, a tool for detecting resource-release omission faults in C code. While many tools exist that detect common patterns in software and detect faults as deviations from those patterns, most suffer from high rates of false positives. Hector takes the novel approach of detecting inconsistencies local to a single function, and was thus able to find over 300 faults in Linux kernel code and other C infrastructure software, with a rate of false positives of only 23%. Improving the robustness of the implementation of Hector was the focus of a 2-year engineer position supported by Inria, with the goal of making Hector publicly available and popularizing its use in the Linux kernel developer community. Hector was the subject of a paper at DSN 2013 that received the DSN 2013 Carter award.

In 2014, Lisong Guo completed a PhD around the notion of an oops (crash dump) in the Linux kernel. This work proposed an algorithm for relating the assembly code found in an oops report, generated by an unknown user using an imprecisely known kernel version, to the kernel's C source code. This work targets particularly the case of a kernel maintainer who is trying to understand crash reports found in an automatically harvested kernel oops repository, as have been maintained at various times by the Linux Foundation or various Linux distributions, to increase the practical value of such repositories.

Our most recent work has focused on driver porting. Specifically, we have considered the problem of porting a Linux device driver across versions, particularly backporting, in which a modern driver needs to be used by a client who, typically for reasons of stability, is not able to update their Linux kernel to the most recent version. When multiple drivers need to be backported, they typically need many common changes, suggesting that Coccinelle could be applicable. With Luis Rodriguez of Suse, we have studied the feasibility of using Coccinelle for various complex and recurrent device driver backporting tasks. Coccinelle is now actively used by the Linux kernel backports project, which maintains backports for a large number of Linux device drivers over time as these drivers evolve.

The next step is to fully automate the backporting (or symmetrically forward porting) process: given any driver for one Linux kernel version, one would like to obtain a driver that has the same functionality for another

kernel version. This requires identifying the changes that are needed, obtaining examples of how to carry these changes out, and inferring from these examples a change that is appropriate for the given driver code. We have carried out a preliminary study in this direction with David Lo of Singapore Management University; this work, published at ICSME 2016, is limited to a port from one version to the next one, in the case where the amount of change required is limited to a single line of code.

More general automation of backporting requires more extensive search for relevant examples. This raises issues of scalability, because the Linux kernel code history is very large, and of expressivity, because we need to be able to express complex patterns to obtain change examples that are most relevant to a particular backporting problem. To this end, we have been adapted the notation used by Coccinelle, which describes how a change should be carried out, into a *patch query language* that allows describing patterns of changes that have been previously performed. The associated tool, Prequel, can find patches that match a particular pattern among several hundred thousand commits, often in tens of seconds. This work, published at Usenix ATC 2017, is supported in part by OSADL, a consortium of companies, mostly in Germany, supporting the use and development of open source software in automation and other industries.

Research axis 2: Domain-specific languages There exists a significant gap between the high-level objectives of infrastructure software and their implementation in low-level, imperative programming languages. To bridge that gap, we propose an approach based on domain-specific languages. By abstracting away boilerplate code, DSLs increase the productivity of systems programmers, and by providing a more declarative language, DSLs reduce the complexity of code, thus the likelihood of bugs. Julia Lawall, Gilles Muller and Pierre-Évariste Dagand have a strong experience in designing domain-specific languages in a systems context such as Devil, Bossa, Zebu, and Filet-o-Fish.

Our current work on DSLs has two complementary goals: (i) the design of a unified framework for developing and composing DSLs, following our work on Filet-o-Fish, and (ii) the design of domain-specific languages for domains where there is a critical need for code correctness, and corresponding methodologies for proving properties of the run-time behavior of the system. To bootstrap our long-term effort in designing safe and composable domain-specific languages, we have initiated two exploratory actions involving a combination of advanced type-theoretic concepts and domain-specific compilation techniques. Both actions are complementary, the first adopts a bottom-up approach – going from low-level artifacts to high-level abstractions – while the second follows a top-down approach – offering a safe translation of high-level guarantees to low-level executable code.

Our first line of inquiry, of which some early results have been published at FLOPS 2016, aims at bridging the formalization gap between low-level, bit-twiddling code and high-level, mathematical abstractions. As such, it provided us with an opportunity to experiment with using an interactive theorem prover to design abstractions in a bottom-up manner. We have developed a library (`ssrbit`, publicly available under an open-source license) for modeling and computing with bit vectors in the Coq proof assistant. Because ease of proving and efficiency in computing are often incompatible objectives, this library offers a two pronged approach by offering an abstract specification for proving and an efficient implementation for computing; we have shown that the latter is correct with respect to the former. Using this model of bit-level operations, we have implemented a bitset library and proved its correctness with respect to the formalization of sets of finite types provided by the Ssreflect library, which is part of the Mathematical Components framework developed at the MSR-Inria joint center. This library thus enables a seamless interaction of sets for computing and sets for proving. This library also supports the trustworthy extraction of bitsets down to OCaml's machine integers: we gained greater confidence in our model by adopting a methodology based on exhaustive testing. This enabled us to implement three bit-twiddling applications in Coq (Bloom filter, n-queens, and the efficient enumeration of all k-combinations of a set), prove their correctness and obtain efficient low-level OCaml code.

Our second line of inquiry is influenced by the realization that domain-specific languages are often treating the symptoms rather than providing a cure. Infrastructure software is often developed in C, which suffers from many semantic kludges and is, as a result, hardly amenable to formal reasoning. Many domain-specific languages are born out of the frustration of being unable to guarantee static properties of one's code: more often than not, the resulting language is little more than a domain-specific variant of Pascal supporting custom static analyses and some form of transliteration to C. To achieve safety and composability, we believe that a more holistic approach is called for, involving not only the design of a domain-specific *syntax* but also of a domain-

specific *semantics*. Concretely, we are exploring the design of *certified domain-specific compilers* that integrate, from the ground up, a denotational and domain-specific semantics as part of the design of a domain-specific language. This vision is illustrated by our work on the safe compilation of Coq programs into secure OCaml code. It combines ideas from gradual typing – through which types are compiled into run-time assertions – and the theory of ornaments – through which Coq datatypes can be related to OCaml datatypes. Within this formal framework, we enable a secure interaction, termed *dependent interoperability*, between correct-by-construction software and un-trusted programs, be it system calls or legacy libraries. To do so, we trade static guarantees for runtime checks, thus allowing OCaml values to be safely coerced to dependently-typed Coq values and, conversely, to expose dependently-typed Coq programs defensively as OCaml programs. Our framework is developed in Coq: it is constructive and verified in the strictest sense of the terms. It thus becomes possible to internalize and hand-tune the extraction of dependently-typed programs to interoperable OCaml programs within Coq itself.

To further explore the realm of domain-specific compilers, we have been involved in the design and implementation of a certified compiler for the Lustre synchronous dataflow language. Synchronous dataflow languages are widely used for the design of embedded systems: they allow a high-level description of the system and naturally lend themselves to a hierarchical design. This on-going work, in collaboration with members of the Parkas team and Gallium team of Inria Paris, formalizes the compilation of a synchronous data-flow language into an imperative sequential language, which is eventually translated to Cminor, one of CompCert's intermediate languages. This project illustrates perfectly our methodological position: the design of synchronous dataflow languages is first governed by semantic considerations (Kahn process networks and the synchrony hypothesis) that are then reified into syntactic artefacts. The implementation of a certified compiler highlights this dependency on semantics, forcing us to give as crisp a semantics as possible for the proof effort to be manageable. This work was published at PLDI 2017.

In terms of DSL design for domains where correctness is critical, our current focus is on process scheduling and multicore architectures. Ten years ago, we developed Bossa, targeting process scheduling on unicore processors, and primarily focusing on the correctness of a scheduling policy with respect to the requirements of the target kernel. At that time, the main use cases were soft real-time applications, such as video playback. Bossa was and still continues to be used in teaching, because the associated verifications allow a student to develop a kernel-level process scheduling policy without the risk of a kernel crash. Today, however, there is again a need for the development of new scheduling policies, now targeting multicore architectures. As identified by recent work in EuroSys 2016, large-scale server applications, having specific resource access properties, can exhibit pathological properties when run with the Linux kernel's various load balancing heuristics. We are working on a new domain-specific language, Ipanema, to allow expressing load balancing properties, and to enable verification of critical scheduling properties such as work-conservation and liveness; for the latter, we are exploring the use of tools such as model checking, in collaboration with Yann-Thierry Mieg at LIP6, and the Leon theorem prover from EPFL. A first version of the language has been designed and we our initial prototype shows performance on a par with the Linux CFS scheduler.

Research axis 3: Managed run-time environments and multicore architectures During the previous evaluation period, we acquired, within Regal, a solid expertise in the development of Managed Runtime Environments (i.e., language virtual machines, such as the Java Virtual Machine) and in multicore systems. This expertise has led us to initiate several collaborations with industry partners, in the form of CIFRE PhD support.

The work on Managed Runtime Environment ended with the creation of Whisper. Still, two PhD students, Koutheir Attouchi and Florian David, graduated on this topic in 2014 and 2015. The multicore systems topic was initiated with the PhD of Jean-Pierre Lozi on efficient locking. Later, we targeted real-time multicore systems with the goal of improving resource usage, through a cooperation with Renault and the PhD of Antoine Blin. Recently, we have started another cooperation on multicore real-time systems for avionics and space with Thales TRT, that is the topic of the PhD of Cédric Courtaud.

The PhD of Koutheir Attouchi was done in cooperation with Orange Labs in the context of Smart Homes. Market players aim to deploy component-based and service-oriented applications from untrusted third party providers on a single OSGi execution environment. This creates the risk of resource abuse by buggy and malicious applications, which raises the need for resource monitoring mechanisms. In a first paper published at CBSE 2014 we proposed an approach to monitor the memory consumed by each tenant, while allowing them

to continue communicating directly to render services. We modified an experimental Java Virtual Machine in order to provide the memory monitoring features for the multi-tenant OSGi environment. Our evaluation of the memory monitoring mechanism on the DaCapo benchmarks showed an overhead below 46%. In a second paper published at DSN 2015, we addressed the problem of stale references to objects of outdated class loaders. We developed Incinerator, a Java Virtual Machine extension, that detects and nullifies stale references during a garbage collection cycle. We evaluated the performance of Incinerator with the DaCapo benchmark and showed that Incinerator has an overhead of at most 3.3%.

The PhD of Florian David was on optimizing large multi-threaded Java server-class applications that use locks to protect access to shared data. Understanding the impact of locks on the performance of a system is complex, and thus the use of locks can impede the progress of threads on configurations that were not anticipated by the developer, during specific phases of the execution. In a paper, published at OOPSLA 2014, we proposed Free Lunch, a new lock profiler for Java application servers, specifically designed to identify, in-vivo, phases where the progress of the threads is impeded by a lock. Free Lunch is designed around a new metric, critical section pressure (CSP), which directly correlates the progress of the threads to each of the locks. Using Free Lunch, we have identified phases of high CSP, which were hidden with other lock profilers, in the distributed Cassandra NoSQL database and in several applications from the DaCapo 9.12, the SPECjvm2008 and the SPECjbb2005 benchmark suites. Our evaluation of Free Lunch showed that its overhead is never greater than 6%, making it suitable for in-vivo use.

The PhD of Jean-Pierre Lozi was on improving the performance locks on large multicore architectures. In a paper published at Usenix ATC 2012, and more recently in an article published in ACM Transactions on Computer Systems (TOCS), we proposed a new locking technique, Remote Core Locking (RCL), that aims to accelerate the execution of critical sections in legacy applications on multicore architectures. RCL is currently one of the most efficient locking technique and the ATC 2012 paper has currently 97 citations on Google scholar. The idea of RCL is to replace lock acquisitions by optimized remote procedure calls to a dedicated server hardware thread. RCL limits the performance collapse observed with other lock algorithms when many threads try to acquire a lock concurrently and removes the need to transfer lock-protected shared data to the hardware thread acquiring the lock because such data can typically remain in the server's cache. Eighteen applications were used to evaluate RCL from standard multicore benchmark suites, such as SPLASH-2 and Phoenix 2. By using RCL instead of Linux POSIX locks, performance is improved by up to 2.5 times on Memcached, and up to 11.6 times on Berkeley DB with the TPC-C client. On a SPARC machine with two Sun Ultrasparc T2+ processors and 128 hardware threads, performance is improved by up to 1.3 times with respect to Solaris POSIX locks on Memcached, and up to 7.9 times on Berkeley DB with the TPC-C client.

In an article published in ACM Transactions on Parallel Computing (TOPC), we investigated the applicability of Software Transactional Memory (STM) for reactive applications in which there is a required response time for certain operations. We proposed supporting such applications by allowing programmers to associate deadlines and QoS requirements with atomic blocks. Based on statistics of past executions, we adjust the execution mode of transactions by decreasing the level of optimism as the deadline approaches. In the presence of concurrent deadlines, we propose different conflict resolution policies. Execution mode switching mechanisms allow meeting multiple deadlines in a consistent manner, with potential QoS degradations being split fairly among several threads as contention increases, and avoiding starvation. Our implementation consists of extensions to a STM runtime that allow gathering statistics and switching execution modes. Our experiments show that our approach significantly improves the likelihood of a transaction meeting its deadline and QoS requirement, even when progress is hampered by conflicts and other concurrent transactions with deadlines.

The PhD of Antoine Blin is on modern complex embedded systems that involve a mix of real-time and best-effort applications. The recent emergence of low-cost multicore processors raises the possibility of running both kinds of applications on a single machine, with virtualization ensuring isolation. Nevertheless, memory contention can introduce other sources of delay, that can lead to missed deadlines. In a paper published at ECRTS 2016, we have presented a combined offline/online memory bandwidth monitoring approach. Our approach estimates and limits the impact of the memory contention incurred by the best-effort applications on the execution time of the real-time application. Using our approach, the system designer can limit the overhead on the real-time application to under 5% of its expected execution time, while still enabling progress of the best-effort applications.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	1	2	3	1	1		8
Ouvrages							
Conferences	2	4	8	5	5	5	29
Logiciels		1			2	1	4
Brevets							

Sélection des produits et des activités de recherche

Faits marquants

- The main highlight of these years is the continuous spreading of Coccinelle within the developer community of the Linux kernel. We submitted the first patches to the Linux kernel based on Coccinelle in 2007. Since then, over 4500 patches have been accepted into the Linux kernel based on the use of Coccinelle, including around 3000 by over 500 developers from outside our research group. Another testimonial of the impact of our work is the signature of a Memorandum Of Understanding (MOU) with the Linux Foundation. As part of the MOU, Greg Kroah-Hartman is spending a year with Whisper since October 2016. Kroah-Hartman is one of the leading developers of the Linux kernel, and is one of only a few developers employed by the Linux Foundation, with another being Linus Torvalds. Greg participated in the activities of the Whisper team around the use of Coccinelle and research projects related to the Linux kernel, and he is a convinced ambassador of our research work.

- We received the best paper award at ASE 2012 for our work on Diagnosys: Automatic Generation of a Debugging Interface to the Linux Kernel.

- Suman Saha received the William C. Carter Award from DSN 2013. The award recognizes an outstanding paper based on a graduate dissertation, and is the only form of best paper award given at DSN. The award was given for the paper Hector : Detecting Resource-Release Omission Faults in Error-Handling Code for Systems Software.

- The paper “Faults in Linux 2.6” was published in the ACM journal Transactions on Computer Systems in June 2014 which is the most prestigious journal in systems. It has been cited more than 170 times according to Google Scholar. The paper was reviewed in the Linux Weekly News, in the German professional IT website golem.de, and was the subject of an invited presentation at a joint session of the Linux Kernel Summit and LinuxCon North America.

- Our work on Remote Core Locking (RCL) (PhD of JP Lozi) was accepted in ACM Transaction in Computer Systems (TOCS) in January 2016. RCL is currently one of the most efficient locks for multicore architectures. The RCL was first published at Usenix ATC in 2012, and that paper has been cited about 100 times according to Google Scholar.

- Our foundational work providing high-level abstractions in DSLs for low-level software has been well received with a best paper award at FLOPS in 2016.

3. Analyse SWOT

About research axis 1, the effort put into the development and continuing maintenance of Coccinelle is fairly unusual in the research community. The degree of maturity of Coccinelle has facilitated an ongoing interaction with the Linux kernel developer community, and the continuing study of Linux kernel code and kernel development issues has inspired new research directions. This does entail a moderate and continuing cost in terms of researcher time, however, we have been aided in recent years by very competent engineers with Inria, who help address issues that are important to users but are not critical to current research. We have recently been less successful in attracting good PhD students in this area, with two students who abandoned their studies at or before the half way point. Our main target for the near future in this area is the further development of Prequel

and of related approaches needed for driver porting. We would also like to explore other aspects of the kernel that may benefit from program analysis, such as device trees.

About research axis 2, our foundational work providing high-level abstractions for low-level software has been well received with a best paper award at FLOPS 2016. It takes advantage of a thriving French school of formal reasoning, providing us with highly qualified students and local expertise in the use of advanced tools such as the Coq proof assistant, which is developed at Inria. We wish for this library to be used at large in the Coq community, where several projects are in dire need of a standard library of bit vectors. Our work on domain-specific compilers is still in its infancy, with our various contributions only beginning to bear fruit, such as our ICFP 2016 and PLDI 2017 papers. So far, our work is purely of academic interest: we wish to expand our collaboration to industrial partners willing to invest in research on the development of safety critical systems. The work around Ipanema started at the end year 2015 and the first scientific results have been reported in our HOTOS 2017 paper. We do not yet have dedicated funding to support the development of Ipanema. Still, Redha Gouicem obtained a PhD fellowship in October 2016 from the Edite doctoral school to work in this area.

About research axis 3, multicore architectures are still an emerging topic and a challenge for many companies in the domain of embedded systems. For Whisper this is an opportunity to initiate cooperations with industry, and to disseminate our knowledge (and our tools). We have found that CIFRE PhD fellowships are a very convenient way to establish a cooperation. Still, the main challenge is to establish a long term cooperation, given the changing strategies of many industry partners.

4. Projet scientifique à cinq ans

We recall that Whisper was created recently. Therefore, our research directions for the four next years are straightforward continuations of the work we started during the first three years of the team.

In the direction of improving legacy infrastructure software, we are interested in exploring a number of areas. Building on the results obtained with Prequel, we expect that a main focus will be the investigation of techniques for porting Linux kernel device drivers between Linux versions. Indeed, the rate of change of the Linux kernel makes it difficult for clients to keep up, due to the need for stability in their products. Older kernels, however, need newly developed services, particularly device drivers, while code developed to target an older kernel needs to updated to the latest mainline release. Many cases of porting between versions simply require finding examples of how to update incompatible usages, e.g., functions that are recently introduced or no longer defined, and then generalizing these examples to form a rule that is correctly applicable to the code that needs to be ported. Our experiments show that Prequel already makes it possible to find examples. The next step is to generalize of these examples into widely applicable rules. This process is challenging, because correctly performing a change may require taking into account context information, and it can be difficult to identify the relevant context in an efficient and useful way. This work is supported by the ITrans PRCI ANR, in collaboration with David Lo an Lingxiao Jiang at Singapore Management University.

At a wider scope, there are many opportunities to use Coccinelle and Prequel, and the methodologies developed in these projects, to address infrastructure software development issues. We are currently interested in exploring the consistency of device tree declarations and the kernel code that depends on them, in the context of a potential collaboration with a subcontractor of Renesas. We are also interested in developing tools to understand where locks are needed; researchers at the University of Dortmund have expressed interest in developing a collaboration in this area. We will continue to put effort into the development and dissemination of Coccinelle, as a vehicle for interacting with the Linux kernel developer community. Finally, we hope to build on the Memorandum Of Understanding (MOU) signed between Inria and the Linux Foundation to enhance our collaboration, both as a way of learning about emerging issues and as a way of increasing the visibility of our results.

In the direction of domain-specific languages, we are exploring two main axes: The first is foundation oriented and is related to exploring the design of *certified domain-specific compilers* that integrate, from the ground up, a denotational and domain-specific semantics as part of the design of a domain-specific language. The second is more practical with the design of real OS services. Our current target is the Ipanema domain-specific language, whose goal is to express policies for allocating tasks to processors in a multicore setting. Here, the main challenge for the next year is to prove properties of the behavior of scheduling policies during execution. In the longer term, we will explore the generalization of Ipanema to other resources such as memory for virtual machines. In this context, we also hope that the MOU with the Linux Foundation will help the

dissemination of our results to the Linux community.

Multiprocessor architectures are an emerging concern for the real-time and embedded industries as exemplified by our cooperations with Renault and Thales Research. Therefore, this will be an application domain of increasing importance for the Whisper project-team through the next evaluation period. This is first an opportunity for establishing cooperations with industry partners by disseminating our expertise in systems and programming languages. It is also a way to propagate our tools.

DEPARTEMENT SOC

1. Présentation du département

Overview and Composition

The *System-On-Chip* (SoC) research department in the LIP6 laboratory brings together researchers working on different aspects of electrical and computer engineering, including smart devices, computer architecture, hardware security, quantum information, embedded systems, and dependable systems.

As of June 2017 the SoC research department consists of 27 Full Professors and Associate Professors, 6 permanents researchers, 26 PhD students, 8 post-docs and research engineers.

History

This research department was created in 2006 as a result of the restructuring of the LIP6 laboratory. In 2006, it was composed of 2 teams, ALSOC and CIAN. In 2009, the SYEL team joined the SoC department, complementing the activities of the ALSOC and CIAN teams and bringing in new activities on transport and E-health applications. In September 2016, the QI team joined the department expanding the scope of the department to include quantum information.

Pr. Alain Greiner was the head of the department between 2006 and 2009. Pr Alix Munier was the head of the department between 2009 and 2013. Marie-Minerve Louërat has been the head of the department since 2013.

Structure

The department is structured in 4 teams:

ALSOC: The team activities concern methods and tools for multiprocessor system on chip design (MPSOC). Such highly integrated multiprocessor architectures are used in embedded applications such as automotive, audio-video, image processing, computer vision and telecommunications. The design of these systems requires the development of hardware and software co-design methods. The team focuses on advanced hardware manycore architectures, communication protocols, embedded operating systems, real-time constraints, formal methods for verification systems, and optimization of design and deployment of algorithms specifically for manycore architectures.

CIAN: The team has a multidisciplinary expertise on topics spanning wide areas of electrical engineering and computer science. It targets challenges resulting from the miniaturization, complexity, and heterogeneity (i.e. multi-domain, multi-physical) of modern integrated circuit technologies and the demand for achieving high dependability (i.e. maintainability, availability, reliability, durability, safety) and security. To meet these challenges, the team focuses on the development of new architectures as well as on methods and tools for modeling, design, test, and verification.

QI: The team works on various aspects of quantum information, quantum communication and quantum computation, from foundations to implementations. This includes the development of cryptographic protocols including quantum key distribution, quantum coin flipping, blind quantum computation, verified quantum computation, secret sharing, and secure multiparty quantum computation. QI tries to understand the fundamental quantum features which give rise to quantum advantages, and how through information perspectives we can understand better the foundations of quantum mechanics.

SYEL: The team studies at the same time fundamental and applied aspects in Computer Engineering at the frontier between Computer Science and Electronics. On one hand, members of the team contribute to the fundamental advances in the field of electronic systems, mainly their modeling and their architecture. On the other hand, they exploit these advances to design dedicated systems for a class of applications. The aim of these research activities is the validation of new architectures for embedded systems and their evaluation in application specific domains, such as transport and E-health. These evaluation procedures generate new constraints and new open questions, which in return feed the activities of the team.

2. Politique scientifique

Scientific goals

The SoC department addresses a broad range of research topics in electrical and computer engineering, ranging from hardware to software:

— Design of computer architectures, tools for profiling and optimizing parallel applications, design of high performance applications for embedded systems.
— Design and test of integrated circuits (ICs) and smart devices deployed in wireless communication systems and autonomous sensor networks: digital processors, signal processing, network of synchronous FPGA, analog, RF, sensors, actuators, and MEMS.
— Design of embedded systems.
— Electronic Design Automation (EDA) methods and tools: Multi-domain modeling (i.e. sensors, actuators, and RF interacting with software), automated design flows, verification, hardware description language (i.e. SystemC AMS), failure analysis, fault diagnosis, test, machine learning applications in EDA.
— Security: Software robustness verification and code hardening, hardware security (i.e. reverse-engineering, side-channel attacks, fault injection, counterfeiting, hardware Trojans).
— Practical systems for quantum cryptography and communication complexity (continuous-variable QKD, quantum money, etc), silicon photonics implementations.
— Secure quantum network protocols (secret sharing, blind computing, secure multiparty computation, verification methods), development of multiparty resources.
— Foundations of quantum mechanics and implications to quantum information.
— Social impact of microelectronics.

The targeted applications include wireless communication systems, sensor networks, E-health, transport (automotive, avionics), etc.

In coming years, the 4 teams will focus on (a) software and hardware co-design in large SoCs; (b) hardware security; (c) verification and post-silicon validation of complex, mixed-signal, heterogeneous SoCs; (d) middleware, verification, optimization and algorithm design for manycore architectures taking into account security constraints and objectives; (e) development of verifiable quantum networks, both in theory and experimental implementation. Synergy is envisioned between the ALSOC, CIAN, and SYEL teams and the new QI team by leveraging the experience that the ALSOC, CIAN, and SYEL teams have accumulated on modelling, design, verification and test of CMOS-based chips towards the efficient and robust implementation of quantum computers.

Dissemination strategy

Start-ups: The department SoC has been involved in the creation of 4 spinoffs during the last period:
— FLEXRAS created in 2009 by Hayder Mrabet (PhD CIAN) and sold in 2015 to Mentor Graphics
— INTENTO Design created in 2015 by Ramy Iskander (Associate Professor CIAN)
— SEAMLESSWAVES created in 2016 by Hassan Aboushady (Associate Professor CIAN)
— WISEBATT created in 2016 by Wilfried Dron (PhD SYEL)

FOSS: The department SoC is also involved in the distribution of Free Open Source Software:
— ALLIANCE: https://www-soc.lip6.fr/en/team-cian/softwares/alliance/
— CORIOLIS: https://www-soc.lip6.fr/en/team-cian/softwares/coriolis/
— OCEANE: https:/www-soc.lip6.fr/en/team-cian/softwares/oceane/

EQUIPE ALSOC

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

ALSOC team is a 11-year old team whose hystorical research is computer architecture at large. For more than 10 years the research focuses on the design of a manycore architecture TSAR (Tera Scale ARchitecture), its network on chip and its operating system and tools for architectures. Now the team has started to shift to more software aspects dealing with architecture.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	3	3	3	4	4	4
Researchers	8	7	8	9	9	9
PhD students	17	17	13	11	12	8
Others	1	1	1	1	1	1
Total	29	28	26	26	28	24

Grants	2012	2013	2014	2015	2016	2017
International						
National	253	255	277	191	85	82
Industry	23	23	41	31	34	12
Others						
Total	276	277	318	222	119	94

Arrivals : Hamed Sheibanyrad (CNRS CR1 researcher) in 2014, Cécile Braunstein (associate Professor) in 2015, after a granted leave as a researcher in Bremen (2012-2015), Lionel Lacassagne (Professor) in 2015.
Departure : Francois Pêcheux (associate Professor) in 2013 (nominated Professor in LIP6/SOC/Cian team).

Politique scientifique

The research done in the ALSOC team is mainly applied research. For the last period, the activities of the ALSOC team concern methods and tools for multiprocessors system on chip design (MPSOC) and their software stack. The team focuses on advanced hardware manycore architecture, communication protocols, embedded operating system, real-time constraints, formal verification of systems, code hardening and optimization, deployment design and optimization of algorithms specially designed for manycore architecture.

2. Produits de la recherche et activités de recherche

Bilan scientifique

For the previous AERES evaluation, three themes were developped by ALSOC. Here are the results, five years after :

Theme #1 : Modeling and simulation of complex systems

The competences on architecture and scheduling were combined in an effort to solve the packet reordering problem, critical in massively parallel networking applications, both on algoritmic and implementation level (JSA2013).

A virtual prototype platform based on Multi-level model transformations and generating a mapping onto SoCLib platforms is now available for design, analysis and verification (model checking) of heterogeneous multi-core systems (MODELSWARD 2017). It is currently used for ADAS (Advanced driver-assistance systems) and control-bound applications. This platform provides a formal semantics based on timed automata for embedded applications written in an extended version of the Kahn process network model developed by the team in the past. This work is a close collaboration with the LabSoC team of Telecom ParisTech.

Theme #2 : Design of many-core architecture with shared memory

The TSAR platform is now mature and has very strong scaling up to 1024 cores for representative applications belonging to linear algebra, signal and image processing for embedded systems. The L1-L2 cache coherence protocol has been improved (ISVLSI 2015), as well as the TLB coherence mechanism, and hardware fault tolerance has been added by the means of router reconfiguration to avoid faulty clusters. The architecture provides an easy and efficient way (shared memory) for programmers (compared to DMA transfers requests for other existing manycore architectures). Some virtualization extensions have been developed and allow the isolated execution of different operating systems on different parts of the architecture (MICPRO 2016).

Theme #3 : Deployment of applications and their adaptability onto many-core architecture

In order to fully exploit the computing power of NUMA many-core architectures, we proposed a dynamic, transparent, portable approach for task and data placement to be deployed in the runtime associated to a task-parallel language named Open-Stream (TACO 2014, PACT 2016). We also developed a tool named Aftermath for performance debugging (ISPASS 2016, IWOMP 2016). It offers a visualization of cross-layers information enabling developers to point out potential sources of performance bottleneck coming from complex interaction between the application, the runtime, the OS and the underlying architecture.

Research work also addresses methods and tools for optimization (e.g. threads and resources allocation) of multi-threaded applications. An important theoretical effort allowed the development of a set of original mathematical tools to evaluate the vivacity and to compute the maximum throughput of communicating systems described by Synchronous-Data Flow Graphs (SDFG). These methods have been extended in some cases successfully to Cyclo-Static DataFlow Graphs (CSDFG) and some important subclasses of generalized timed Petri nets. We have also developed an original algorithm of low complexity to evaluate the mapping of a SDFG/CSDFG on a many-core machine based on the structure of the graph.

The team started to work on fault attack models at source-code and assembly code levels, formal methods and tools for robustness analysis of assembly code and code hardening of embedded systems subject to physical attacks. From experimental campaigns we studied the modeling of fault models (FDTC 2013, HOST 2014, N. Moro PhD thesis). We also designed generic and formally-verified software protection schemes both at source and assembly level (JCEN 2014, ESORICS 2014, N. Moro PhD Thesis). We also proposed an approach based on formal verification in order to assess the robustness of assembly code subject to fault attacks (CARDIS 2015).

We also started to design efficient parallel (that is fast *and* scalable) algorithm. Concerning linear algebra, we designed a library tuned for tiny matrices used by CERN for tracking particles in realtime (on the fly processing) and for us for motion detection and multi-target tracking. This library is written for SIMD multi-core processors (instanced for IBM Altivec, Intel SSE, AVX, AVX-512 and ARM Neon). It addresses 16-bit, 32-bit and 64-bit computations.

Concerning image processing, we have designed with LRI and MICS a connected component labeling and

analysis algorithm named LSL that is the only algorithm to scale on a multi-core multi-processor architecture (ICIP, JRTIP).

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	1	2	7	1	7	2	20
Ouvrages							
Conferences	15	11	11	6	13	2	58
Logiciels	1	3	3	1	3		11
Brevets						1	1

Sélection des produits et des activités de recherche

— **A 16-core TSAR** (Terascale Architecture) ASIC has been founded by STM and CEA LETI, and makes the validation and the deployment of applications a reality. A 96-core version is under fabrication.

— **Almos** & **GIET-VM** : OSes for TSAR architecture.

— **Aftermath** is a graphical tool for performance analysis and debugging of parallel programs and runtime systems. `http://www.openstream.info/aftermath`

— **Turbine** is a powerful generator of random live Synchronous Data Flow Graphs and Cyclo-Static Data Flow Graphs, that was developed to test the scalability of all the optimization methods.

— **Parallel Light Speed Labeling** (for binary image labeling and analysis) : LSL is the only connected component labeling algorithm that scales on multi-core / multi-processor architecture.

Faits marquants

The **TSAR** project led by Professor Alain Greiner, two Associate Professors and ten PhD student during the last 5 years has become a reality. It is a multi-cluster many-core architecture with a user-friendly API (threads + with cache-coherency), that strongly scales up to 1024 cores. There is now a full eco-system with specific OSes like ALOS-MK and GIET-VM (for handling Virtual Memory) but also general purpose OSes like BSD and Linux. It is programmable in C with GCC toolchain. A set of representative applications/benchmark and the simulation environnement based on SocLib (cycle accurate bit accurate) prototyping tool are available to download.

International collaborations : CERN, APT(U. Manchester), AGRA & DLR(U. Bremen), CSL(U. Ghent).

In the context of CERN/LHCb experiment, our SIMD implementation of Choleski factorization for tiny matrices (3×3 up to 16×16 outperforms existing libraries (Intel MKL, Eigen) by a factor greater than $\times 10$.

National collaborations : LRI & LIMSI (Univ. Paris-Sud), MICS (CentraleSupelec), IETR, INSA Val-de-Loire, Lab-Sticc (U. Bretagne Sud), LHC(U. Saint Etienne), LIG(U. Grenoble), CEA LIST, PARKAS (INRIA/ENS), ST Microelectronics, Telecom (LabSoC, LTCI).

3. Analyse SWOT

Strength :

— The team gathers members with strong and complementary expertises in various domains : computer architecture design, compilation & code optimization, task & SIMD parallelism, formal methods, operational research.

— All the activities are pulled by the industry and rely on real needs.

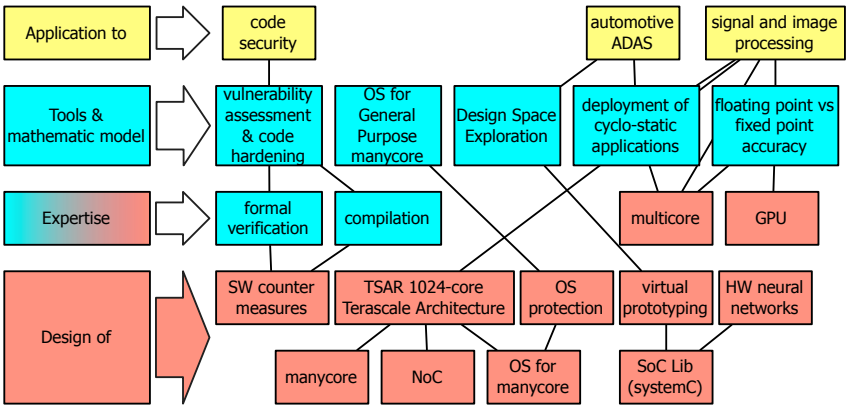


FIGURE 1.1 – ALSOC research areas and connexions

Weakness :

— architecture design requires *long term efforts* and funding,
— several people has heavy investment un teaching missions,
— few international collaboration and european projects.

Opportunities :

— Rise of manycore era but lack of tools and algorithms for programming such architectures,
— TSAR is fully operational : a 16-core prototype is fabricated and will be used instead of the simulator for upcoming parallel implementations,
— security is a hot topic,
— LIP6 complementary expertises in various fields.

Threats :

— decreasing dynamics, fundings and market for multi/manycores hardware design in France and Europe,
— Security is a highly competitive domain,

4. Projet scientifique à cinq ans

EFFICIENT EMBEDDED SYSTEMS - SPEED, SECURITY AND ENERGY EFFICIENCY

ALSOC is at a turning point with a shift to more software aspects of it research dealing with architecture. ALSOC's unique experience in architecture will hep to design new tools, algorithms and methods in order to capitalize on its expertise and move towards new challenges. This will be achieved through software release, publications conferences and journals, national and European funding as well as unfunded collaborations where research is very active and limitless.

1. **Architectures : design of parallel architectures and their OS**

   — Manycore architecture for the design and the evaluation of parallel algorithm : TSAR will be now used to deploy parallel algorithms and to evaluate middlewares.

   — Manycore OS design mixing the multi-kernel approach with direct inter-kernel accesses to improve performance (evolution of Almos-MK).

   — Virtual Prototyping and Design Space Exploration : propose a complete Design Space Exploration Tool based on Model Transformations between multiple levels of representation starting from SysML and arriving at cycle accurate bit accurate level.

2. **Optimisations : methods, algorithms and tools to optimize the performance of embedded systems :**

— Deployment of real-time applications : we proved formally that the communications of a real-time application or any application specified using a Simulink formalism can be modeled using particular Synchronous-Data Flow graphs (SDFG). These specifications can be used to develop new original algorithme to optimize the execution of these applications on a many-core architecture, following as example the AUTOSAR standard.

— Floating-point *versus* fixed-point coding : to develop tools and applications with the required amount of accuracy and to evaluate some trade-off (speed *vs* accuracy *vs* power consumption) applied to computer vision for smart cameras, robots and UAVs.

— Design of efficient irregular parallel algorithms : focus on Connected Component Labeling and Analysis algorithm – an important class of algorithms in image processing, – that combines irregular graph processing (Union-Find building and update), discrete geometry and data-dependent (features) computations to find out how to transform, re-think and design irregular parallel algorithms, for multi/many cores processor with SIMD extension and GPUs.

3. **Code hardening and security : methods, algorithms and tools.**

— The proposed formal verification approach and associated tool for the assessment of robustness of an application subject to fault attacks will be extended to enable an automatic analysis of sensitive parts of an application. We aim at extend the set of considered attacks (fault model and side-channel attacks), the type of security properties as well as the ability of our tool to deal with more complex codes. We aim at investigating the combination with other formal approaches in order to improve the scope of our framework from high-level security requirements to low-level robustness analysis.

— Compilation flow for code hardening and robustness analysis of code subject to physical attacks. A security-oriented compilation flow will be set up in order to produce necessary and useful information for the robustness analysis. We also want to develop hardening pass and study the interaction between securing and optimization pass. This point is partially addressed in a collaborative project and a PhD thesis (ANR PROSECCO and CIFRE ISSM).

— The secure execution of different operating systems on the same architecture requires a secure boot authentication mechanism in order to guarantee the code authenticity and form a chain of trust, from the start of the boot-loader to the kernel. This aspect has not been treated in the TSUNAMY project and may be subject to future work.

The two first axes of code hardening and security theme partially rely on fundings from ANR, 3 phD thesis have recently started and this theme gather 4 permanent staff with complementary skills (formal methods, hardware architecture and OS, compilation).

EQUIPE CIAN

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The CIAN team consists of 11 permanent researchers (Associate, Emeritus, and Full Professors, as well as CNRS researchers), 7 PhD students, 2 post-docs, and 2 research engineers. The team has expertise in electrical engineering and computer science. It aims to provide solutions to the challenges resulting from the miniaturization and heterogeneity of integrated circuit technologies and the need for high reliability and security. Miniaturization leads to the explosion of complexity in terms of components and to the rise of defect levels. The group also focuses on the coexistence of multi-domain and multi-physical environments (analog/RF, digital, MEMS, magnetic, organic. etc.).

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	7	6	8	8	7	7
Researchers	1	1	1	1	2	2
PhD students	22	24	21	17	14	7
Others	5	5	8	7	4	4
Total	35	36	38	33	27	20

Grants KEuros	2012	2013	2014	2015	2016	2017
International	400	518	468	386	102	44
National	277	151	206	149	108	196
Industry	31	64	45	19	25	9
Others	64	64	16			
Total	773	796	735	554	234	250

Politique scientifique

Team members have expertise in various areas of electrical engineering and computer science and are active on topics that fit very well into the core research orientations of LIP6, including smart devices, security and reliability. Most of the projects carried out are application driven and all projects culminate with a demonstrator, namely a circuit prototype or a computer-aided design tool, that serves as a proof of concept. The team strongly encourages technology transfer by communicating the projects' results to interested specializing companies, as well as through the creation of start-ups. The team seeks establishing collaborative links with other teams in the laboratory, as well as with academic and industrial partners form France and abroad, by means of national and international projects. Team members contribute to the organization of conferences and serve as program committee members of conferences and editors in international journals.They also participate very actively in reviewing projects submitted to national and international funded research programs.

2. Produits de la recherche et activités de recherche

Bilan scientifique

The research topics covered by the team can be grouped in five axes :

Hardware Security The team addresses the problem of hardware security by proposing hardware solutions and associated CAD tools and methodologies. Specifically, we are focusing on the following aspects : (a) Development of a design methodology for prototyping efficiently and rapidly digital cryptographic algorithms on hardware, taking into account several metrics, such as frequency, area, etc. This methodology is based on a large library of cryptographic basic blocks. Specifically, using this methodology, we proposed a hardware implementation of AEGIS which participates in the CAESAR competition organized by the NIST. We also proposed several implementations of Authenticated Encryption and Elliptic Curve Cryptography depending on throughput, frequency, and area constraints ; (b) Side-channel attacks to measure the vulnerability of our architectures or propose verification solutions aiming at establishing efficient countermeasures ; (c) Development of a secure FPGA using WDDL logic and the associated design flow, including a simulator-based power analyzer ; (d) Development of a portfolio of obfuscation methodologies for equipping analog ICs with a capability that prevents reverse engineering and counterfeiting ; (e) Development of spectrum sensing Radio Frequency (RF) circuits that are used to sweep the spectrum and detect malicious RF attacks.

Integrated Systems Dependability The team's research activities on dependable circuits and systems include : (a) development of innovative algorithms for fully synchronizing a distributed generation of clocks into large SoCs ; (b) analysis of aging and electromigration phenomena towards design-for-reliability ; (c) diagnosis and failure analysis of mixed analog-digital SoCs towards improving the manufacturing process and boosting yield, facilitating silicon debugging, and identifying the root cause of failures occurring in the field of operation so as to improve safety features in future product generations ; (d) design-for-test and built-in self-test for mixed-signal circuits with the objective of reducing manufacturing test costs and improving outgoing quality ; (e) reusing the existing built-in self-test infrastructure in the field of operation, especially in the context of safety-critical and mission-critical applications, with the aim to perform on-line tests in idle times or concurrently with the operation ; (e) self-calibration and self-healing for mixed-signal circuits and systems aiming at recovering yield loss in post-manufacturing and at achieving fault tolerance and adaptation to the environment in the field of operation.

Design and CAD tools for heterogenous systems The team's research activities on Design and CAD tools for heterogenous systems include : (a) Designer-centric analog circuit synthesis flow. This CAD flow is fully controlled by the designer and offers an intuitive design approach. It captures the designer knowledge to automatically generate the analog IP design procedure and physical view, ensuring consistency and accuracy. Using both simulation-based and knowledge-based approaches, and addressing both sizing and layout generation, the flow has successfully been used to synthesize transconductors and state of the art analog to digital converters (ADCs) ; (b) Low-power RF IPs. Based on its strong experience in the domains of sigma-delta modulation and RF ICs, CIAN has developed an innovative design methodology dedicated to the implementation of state-of-the-art low-power multi-standard RF transceivers capable of dynamically detecting and communicating with surrounding wireless sensors, using the most appropriate communication standard. This requires to specify with great care the RF transceiver operating limits and to control the different operating modes with efficient dedicated reconfigurable digital logic ; (c) Energy harvesting and power management at micro-scale level. Internet of Things (IoTs) devices still suffer from energy loss, and this prevails the massive development of large scale networks of communicating heterogeneous systems, composed of sensors, a processing unit, and a RF transceiver. The solution to this issue is to do on-chip energy harvesting so as to optimize energy sources and to efficiently manage the available system energy. Since 2007, CIAN has developed methods, IPs, and interfaces to handle micro-scale energy sources and to soundly integrate them into low-power SoCs. A noticeable result is the operational design of a MEMS electrostatic vibration energy harvester ; (d) Virtual prototyping of heterogeneous systems at high level of abstraction. Since 2007, and as a member of the Accellera Systems Initiative, CIAN has participated in the specification, the Language Rule Manual (LRM) and the user's guide for the analog mixed signal extensions of SystemC. CIAN has developed a strong experience in the modeling and simulation of heterogeneous systems at a high level of abstraction. This allowed the design of a unified yet flexible multi-disciplinary virtual prototyping environment based on SystemC. This environment allows the simulation of a complex heterogeneous system as a whole, for which each component (whatever its related domain is, i.e. digital, analog, RF,

optical, fluidic) is described and solved using the most appropriate Model of Computation (MoC).

Reconfigurable Architectures The team works on the internal architectures of FPGAs and especially the interconnection networks of tree-like. The first objective is to develop a high-performance architecture for embedded FPGA in a SoC (eFPGA). We have proposed a mixed network architecture based on mesh and tree (Tree of Mesh). We also seek to optimize the architecture for implementing specific applications. Studies have been conducted to develop tree architectures in 3D. All studies were made in hardware and software by developing architectures and associated tools (placer, router, partitioner etc.). The current orientation concerns the configuration flows for multi-FPGA systems. The team has also designed and implemented reconfigurable RF receivers that can adapt to several wireless communication standards operating at different frequency bands.

Microelectronics & Society Microelectronics, a fundamental technology underlying much of the digital world, has important social ramifications. Unlike other research groups in microelectronics, the CIAN team examines the interactions between digital technologies and social groups. An important theme is the study of microelectronics innovation, notably Moore's Law. The group researches the impact of microelectronics and computer manufacturing on occupational health and environmental pollution. It also examines the intellectual history of concepts underlying digital technics, such as the history of computer logics and the history of specific devices such as field-effect and bipolar transistors. The research team is also interested in the design of "transparent" design technologies. Computer security is critical to many aspects of our society. Security depends on secure-software as well as secure hardware. The behavior of silicon chips may be altered by small modifications of code that does the circuit synthesis. Using Free Open Source Software (FOSS) to synthesize the full chip layout gives a complete insight of the designer's intents and allows detecting further malicious modifications. FOSS is also interesting to allow low-cost chip designs fabrication for all. This motivates the efforts that the team puts in FOSS CAD for VLSI. On the digital side, the placement and routing have been improved, based on new algorithms (Coriolis project). On the analog side, we have taken advantage of the digital improvements to work on a Place and Route tool for analog and mixed-signal circuits whose importance is growing in the IoT era since they are used for receiving and transmitting data and for interfacing physical data to the digital processor.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	5	9	10	15	15	2	57
Ouvrages	1	0	0	1	2	0	4
Conferences	16	38	38	19	20	0	131
Logiciels	4	updates	updates	updates	updates	updates	4 + updates
Brevets	2	2	1		2		7

Sélection des produits et des activités de recherche

The selection is provided in "Annexe 4", with a dedicated section related to "Faits marquants".

Faits marquants

1. Proof of concept in Silicon : The team is capable of implementing solutions all the way form conceptual modeling, to design, to final physical implementation. By the end of most projects, we demonstrate circuit prototypes in advanced technology nodes. Examples include : a) the design of high-performance RF communications for IoT [**CIAN-FM-3**], b) the design of a RISC processor for integration into the ALSOC TSAR manycores project, c) the development of an innovative solution for the synchronization of clocks based on a parallel PLL network [**CIAN-FM-1**], and d) a FPGA architecture secured against DPA attacks [**CIAN-FM-2**].
2. The team has launched several start-ups in the recent years (FlexRas focusing on FPGA partitioning, Intento Design focusing on automatic analog circuit synthesis and optimization, and Seamless Waves

on cognitive radio design). FlexRas was recently acquired by Mentor Graphics, one of the top three firms on EDA.

3. The team has released Free Open Source Software :

   CORIOLIS for placement and routing of CMOS circuits : www-soc.lip6.fr/en/team-cian/softwares/coriolis/

   ALLIANCE for VLSI CAD flow, including a library of standard cells : www-soc.lip6.fr/en/team-cian/softwares/alliance/

   OCEANE for sizing analog functions integrated in silicon : www-soc.lip6.fr/en/team-cian/softwares/oceane/

   TAS for timing analysis : www-soc.lip6.fr/en/team-cian/softwares/tasyagle/

   These tools are used in other universities for teaching. In the past 5 years, the CIAN team was contacted by 3 research institutes (IMEC in Belgium, ETH in Switzerland and Tokai University in Japan) that wanted to use our FOSS solution for integrated circuit design, after having encountered issues related to publication, dissemination and reuse of their own designs developed with industrial CAD tools. The CORIOLIS platform, especially, has been selected to participate in the submission of a H2020 proposal to provide a FOSS solution to the design of photonics chips.

4. Conference Chair and organization
   - Marie-Minerve Louërat was General Chair of the 2013 Forum on specifications and Design Languages (FDL) [**CIAN-E-2**]
   - Haralampos Stratigopoulos was Program Chair of the 2015 IEEE International Mixed-Signal Testing Workshop (IMSTW)
   - Haralampos Stratigopoulos was Program Chair of the 2017 IEEE European Test Symposium (ETS)

3. Analyse SWOT

- *Strengths ;* In the past 5 years, CIAN members have :

- Published more than 200 peer-reviewed papers at major international journals and conferences. Journals include IEEE TCAD, IEEE TCAS-I, IEEE TCAS-II, IEEE Design & Test, Springer-JAICSP, Elsevier Integration, the VLSI journal, JMM, IJBC, APL, etc. Conferences include Reconfig, FPT, DAC, ICCAD, DATE, ITC, VTS, ETS, CICC, RFIC, ISCAS, PowerMEMS, MEMS, etc.
- Published more than 5 books or book chapters.
- Received more than 3.3 M euro sponsored research projects from the European Commission, French National Research Agencies (ANR, MINEFI, BPI-France), and industry (STMicroelectronics, ams AG, NXP, Valeo, Continental, etc.).
- Received several awards (French ministry of research iLAB, Best Paper Award 2015 IEEE ETS, Eugene S. Ferguson Prize of the Society for the History of Technology etc.).
- Given full-day tutorials at conferences (DATE, ITC, ISCAS), presentations in special sessions and panels (IMSTW, RFIC, NewCAS, ISVLSI, VTS, ETS, ECCTD), and talks at Collège de France.
- Served as Associate Editors for several international journals (IEEE TCAS-II, IEEE TCAD, IEEE Design & Test, etc.).
- Served in the TPC of several conferences (DATE, ICCAD, ITC, VTS, ETS, etc.).
- Served in the OC of several conferences (General Chair IMSTW'15, Program Chair ETS'17, Topic Chair DATE'13-'15, Track Co-Chair ISCAS'17, etc.).
- Participated in standardization groups (Accellera SystemC Systems Initiative AMS working group).
- Cooperation with academia outside France : University College Dublin (Ireland), The University of Texas at Dallas (USA), Polytechnic University of Catalonia (Spain), Carnegie Mellon (USA), University of Athens (Greece), French University in Egypt (Egypt), Federal University of Santa Catarina (Brazil), Monterrey Institute of Technology and Higher Education (Mexico), Graduate Research Institute for Policy Studies (Japan) etc.

- *Weaknesses ;* The facility at LIP6 for chip measurement and characterization is under-equipped especially for high frequency and RF integrated circuits.
The size of the team tends to decrease because of various budget cuts by the government which leads to a reduction of the number of selected projects.

- *Opportunities ;* Recruitment of new colleagues :

— The CIAN team had the opportunity to recruit Professor Christophe Lécuyer who has a great academic record and industrial experience in the USA and is a specialist of the history of Silicon Valley. He worked on Moore's law and the evolution of semiconductor technology. His policy advice is an asset when brainstorming for research perspectives.

— Amine Rhouni, CNRS Research Engineer in the Electronics and Computer Science domain, has joined LIP6 in December 2016, and can contribute in the development of circuit and systems prototypes to provide strong proofs of concepts.

- *Threats ;* CIAN developed several CAD tools (Alliance, Coriolis,...) and made them publicly available with free licenses. Unfortunately, CIAN does not always have the necessary manpower to maintain and update these tools with more features and provide support and debugging to users from academia and industry.

4. Projet scientifique à cinq ans

The advent of emerging technologies such as 3D, the Internet of Things (IoT) and more generally cyber physical systems (CPS) are promising directions allowing the team to leverage its expertise in various fields, particularly the modeling, design, test, verification, reliability, and security of these systems.
Synergy is envisioned between the CIAN team and the Quantum Informatics team that recently joined the SoC department at LIP6 on the verification of quantum devices.
New projects, will allow us to grow and reinforce our research activities, have been submitted and will be funded (2 accepted ANR, 1 accepted Penta, 1 H2020 under review). Projects have been submitted on analog hardware security ; development of built-in self-test (BIST) infrastructure in SoCs and reuse of BIST for the purpose of on-line test and concurrent error detection ; failure analysis and fault diagnosis in automotive SoCs ; on-Chip energy harvesting ; reconfigurable RF transceivers for 5G communication ; secured wireless communication systems for autonomous vehicles ; FOSS for CAD with technologies beyond CMOS.

EQUIPE QI

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The QI team in LIP6 was formed in September 2016, when its three permanent members (all CNRS) moved from neighbouring LTCI (Télécom ParisTech), where they worked together as part of a 5 strong team on quantum information since 2008/9. Our team is composed of the following three permanents, with complementary, interdisciplinary backgrounds :
Eleni Diamanti, PhD (experimental physics) Stanford 2006, CNRS since 2009
Elham Kashefi, PhD (theoretical computer science) Imperial College 2003, CNRS since 2007
Damian Markham [responsable], PhD (theoretical physics) Imperial College 2004, CNRS since 2008.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	-	-	-	-	0	0
Researchers	-	-	-	-	3	3
PhD students	-	-	-	-	5	5
Others	-	-	-	-	3	5
Total	-	-	-	-	13	13

Grants	2012	2013	2014	2015	2016	2017
International	-	-	-	-	520	600
National	-	-	-	-	80	115
Industry	-	-	-	-	0	0
Others	-	-	-	-	0	0
Total	-	-	-	-	600	715

Politique scientifique

We have built a group working on the development and implementation of quantum information, in particular for networks. Our team is comprised of a unique combination of theory and experiment, which allows us to develop protocols from conception to implementation. Our understanding of quantum features and their application to quantum information has allowed us to both develop the foundations of computer science and physics and develop new protocols for the exciting paradigm of quantum networks offering new functionalities and perspectives. Our protocols have been carried out both in the experiments run in our laboratory, and with world leading collaborators outside (e.g. Bristol, Oxford and Vienna).

2. Produits de la recherche et activités de recherche

Bilan scientifique

Quantum information offers the possibility of unparalleled advantages in levels of security and computational power, and quantum technologies promise revolutionary boosts for sensing and imaging. By now quantum communication protocols between two parties are well developed, with systems for quantum key distribution

commercially available and devices for quantum sensing imaging are on the point of being industrialised. Between these and the elusive full-scale quantum computer there is a vast unutilised gap, where our vision of a quantum network sits. We envisige a connected network of quantum and classical devices of varying size, power and applications.

Our research focuses the development, understanding and implementation of quantum networks with the ultimate goal of demonstrating unambiguous certifiable quantum advantage in real world situations and applications.

Development of quantum networks
We work in two main directions on the development of quantum networks. Firstly exploring and developing new protocols and utilities, including quantum money, quantum secret sharing, quantum sensing and quantum communication complexity. A second direction is in the delegation, verification and certification of quantum devices. This question is crystalised for quantum computation, where its utility is precisely what makes it hard to verify and certify - if a quantum computer is not simulatable on a classical computer, then it posses a great challenge to verification and certification. We have pioneered key protocols for this crucial question, which we apply across the spectum of quantum applications.
Understanding advantages in quantum networks
We focus on two questions. Firstly, where does the quantum advantage come from (studying notions such as entanglement, contextuality and steering). Secondly, how can we use these quantum features optimally. These answers feed into our two other themes : this understanding allows us to exploit better the quantum features (development) and to understand which features should be most protected (implementation)
Implementing quantum networks
We implement quantum protocols both in our laboratory and with international collaborators. The exprimental efforts at LIP6 are still being set up (the time has been long partly due to the building works needed) and currently the expriments remain at Telecom ParisTech. We work on integrated and bulk quantum optics and have demonstrated many protocols we invented including quantum money, quantum key distribution and verification of entanglement. Different protocols require different experimental set up, leading us to establish collaborations externally also for implementation.

These three directions link to each other and build on each other forming the basis of out team collaborations.

A key to the success of our approach, given the particularly interdisciplinary nature of the field is the development and nurturing of a community of researchers, both nationally and internationally. We have established a strong, diverse local community in Paris through firstly the QuPa meeting series, and more recently as founding parties of the Paris Centre for Quantum Computing.

We are also in the Steering Committee of both regional (Ile de France DIM SIRTEQ) and national (GDR IQFA) projects on Quantum Technologies, where we lead the Quantum Communications axis, and are members of the Scientific Council commissioned by the French Ministry of Research and Higher Education to structure the French participation to the Quantum Technologies EU Flagship program.

Internationally we are involved in several collaborations across Europe, North America and Asia. We have a particularly strong link with the University of Edinburgh where we have a long history of scientific and teaching exchanges and are in the process of establishing a LIA with the informatics department. We are also a partner at the Networked Quantum Information hub lead by Oxford as part of UK 270M investment in quantum technology. We have ongoing student and researcher exchanges with Tokyo, London and Bristol. Finally, we also have strong industrial links with established collaborations with IDQuantique, Nokia Bell Labs and Thales Alenia Space.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	-	-	-	-	6	9	15
Ouvrages	-	-	-	-	0	0	0
Conferences	-	-	-	-	0	2	2
Logiciels	-	-	-	-	0	0	0
Brevets	-	-	-	-	0	0	0

Sélection des produits et des activités de recherche

Since joining LIP6 in September 2016 our main focus has been firstly to establish ourselves and our research activities, and secondly to build on our community and give international visibility to our work.

— Installed the QI team, swelled in numbers (post docs, PhD students and interns) and set up joint projects with PolSys team at LIP6 and with LKB.
— In order to support our continued momentum we have submitted several projects :
- 3 QuantEra that have passed to the second round (total budget 4 600 000 euro, of which LIP6 1 000 000 euro)
- 2 ANR that have passed to the second round (total budget 800 000, of which LIP6 230 000 euro)
- 3 ANR international (Singapore, Canada, Germany) (total budget 1300 000, of which LIP6 745 000 euro)
- H2020 Cybersecurity call (PCQC budget 300 000 euro)

Faits marquants

— Eleni Diamanti has been awarded the prestigious ERC fellowship.
— We organised two major international conferences (TQC 2017, TYQI 2017) in order to highlight and bring visibility to the new quantum activities in LIP6 as well as serve the community and stimulate collaborations including through hosting several visitors for short and long terms.

3. Analyse SWOT

Strengths
— Dynamic (9 projects, > 150 publications)
— Leadership (Paris Centre for Quantum Computing (Vice Director) ; UK NQIT (Oxford) Quantum Hub (Associate Director))
— International
— Interdisciplinary (Theoretical computer science, physics, experimental quantum optics)

Weaknesses
— Young and small group
— No teaching staff (misses link to teaching part of the university)

Oppertunities
— Paris Centre for Quantum Computation
— Integration into LIP6 and UPMC
— Engagement with industry
— EU Flagship on Quantum Technologies

Threats
— Difficulty in convincing classical cryptography community / industry of quantum usefulness
— Difficulty in integrating with current network technology

4. Projet scientifique à cinq ans

Our general scientific goal is the development of verifiable quantum networks, both in theory and experimental implementation. We will develop quantum cryptographic protocols including quantum money, communication complexity, secret sharing, blind quantum computing, secure multiparty computation and general verification methods. We will push the understanding of the role of foundational quantum features in protocols. We will utilise both classical and quantum cryptographic techniques to provide enhanced network security. We will pursue the experimental implementation of protocols in photonics on and off chips. Another important goal is the integration of our new team within LIP6 and UPMC. We will develop and build on existing projects with other departments in UPMC, notably with the quantum information teams in the Physics Department/ Laboratoire Kastler Brossel (LKB). All of this will be achieved also by developing collaborations in the quantum community, locally through PCQC and internationally through continued projects and notably with the upcoming European quantum technologies flagship.

In addition we have two large developments in the very near future. Firstly we will be installing an ATOS funded chair on quantum information across LIP6 and LKB, headed by Elham Kashefi on the LIP6 side and Nicolas Treps on the LKB side. This offers a great opportunity, not only to pursue our research agenda in an optimal setting, but also engage with industry and bring with it added visibility. Secondly, as mentioned above, we are involved in a submitted EUR teaching program on quantum information, again together with LKB, also involving other members of PCQC at UP7. This program represents a great opportunity to train and recruit top level researchers into the field. It also highlights again the need for a teaching position in our team.

EQUIPE SYEL

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

Members of the SyeL team study at the same time fundamental and applied aspects. It is a Computer Engineering team at the frontier between Computer Science and Electronic. On the one hand, its members contribute to fundamental advances in the field of electronic systems, mainly their modeling and their architecture design. On the other hand, they exploit these advances to design dedicated systems for a class of applications. The aim of these research activities is the validation, by a proof of concept approach, of new architectures for embedded systems and their evaluation in application fields like transport, biomedical and e-health. These evaluations bring light to the system boundaries and reveal its limits for the considered applications in terms of non-functional constraints. These limits and the way to overcome them provide new questions that feed our research.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	2	2	2	2	2	2
Researchers	6	6	6	6	6	6
PhD students	5	9	12	9	8	7
Others				3	7	1
Total	13	17	20	20	23	16

Grants	2012	2013	2014	2015	2016	2017
International International					255k€ EIT-Digital	
National	232k€ Winocod	331k€ SmartEEG				
Industry		297k€ Corac, CIRA	45k€ CIRA	55k€ Air-Liquide		130k€ MinDef
Others			60k€ AEP	185k€ DAMPEC LUTECH	30k€ Labex Smart	2k€ DAMPEC
Total	232k€	628k€	105k€	240k€	285k€	132k€

Politique scientifique

In the SyeL team, research is driven by a methodology, described in figure 1.1, that enables a strong interaction between applications and scientific issues. Based on the application constraints, we define a quasi-optimum electronic architecture and we design it. Afterwards, performances are experimentally validated from a variety of tests and measurements. Our electronic architecture also gives new opportunities in different application fields by pushing back the limits of current electronic systems in use in these other fields. To enhance our research, we also define electronic design methods based on heterogeneous modeling of embedded systems.

Figure 1.2 describes how our methodology is implemented around three majors themes :

— Reconfigurable Architectures,

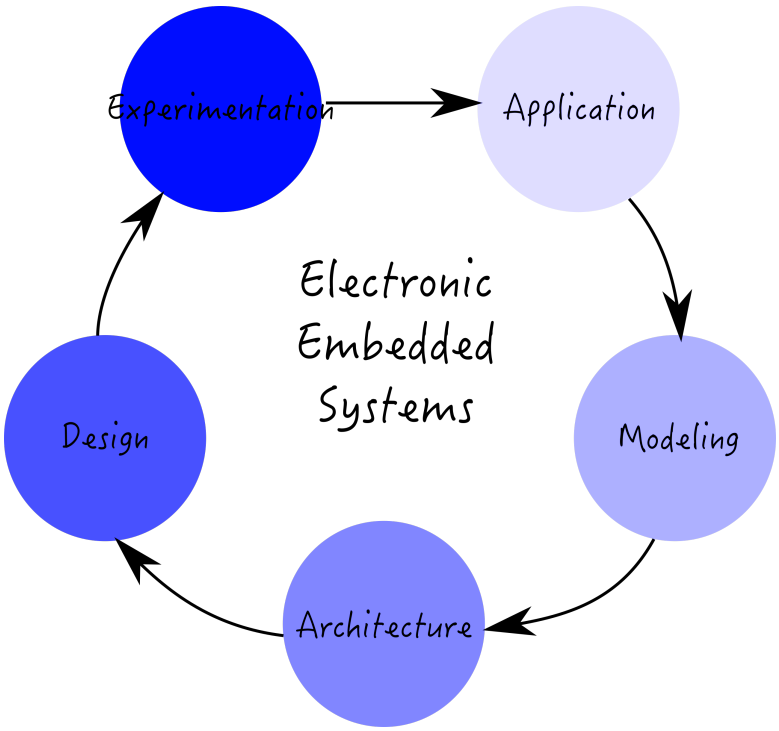


FIGURE 1.1 – SyeL research methodology

— Heterogeneous Modeling,
— Societal Challenge : Biomedical/e-Health and Transport.

Our research activities are conducted in close collaboration between SyeL members team, but also with the other LIP6 team members (CIAN, MLIA, NPA, ACASA, LFI, ...), other French (L2E, INL, ETIS, LabSticc, APHP,...) and international (EPFL, UTH, ULB, RIE, ULPGC, Tokyo, ...) laboratories and with industrials (Thalès, Air Liquide, CIRA, ...).

Reconfigurable Architectures In this theme we study reconfigurable architectures that can be used in embedded systems and that can satisfy non-functional constraints. In the WiNoCoD ANR project (2012-2016), [**SYEL-J-9**], we studied a Radio Frequency network on chip (RF-NoC) that features a reconfigurable OFDMA technique to optimize data transfer latency and processing fill factor for a MultiProcessor-System-on-Chip (MP-SoC). We have built a SystemC demonstrator of a many-core using the RF-NoC.

Also, currently we define and design dynamic reconfigurable architecture. In this context, we have developed a tool and an architecture, named MATIP [1]. With MATIP a user can develop hardware tasks with a VHDL version of the Message Passing Interface library (MPI) and then provide communication facilities for it. It allows for a hardware task the capacity to dynamically create a new hardware task [**SYEL-C-18**].

Heterogeneous Modeling Embedded systems are heterogeneous since they mix analog and digital parts. It is then fundamental to modelize both aspects in a unique model. To do such a task, we work on the influence of technological and geometrical parameters for the analog part. We realize transistor models to study the effect, for example, of W and L on the device sensitivity. We implement energy harvesting models (PEPS CNRS RETHREC, Mita-Shibata laboratory of Tokyo University) and photo-detector modeling, [**SYEL-J-11**].

One of our originality is to use neural networks to modelize the analog part. By using them we have designed a methodology of signal integrity modeling, [**SYEL-J-12**], and the modeling of the charge capacity of batteries. This work is useful to estimate the autonomy of an embedded system, and to manage the system depending on the available power. We have integrated this estimation method in the Contiky OS.

In the HODISS ANR project (2008-12), and its follow-up, we have studied a Digital Ring Oscillator (DCO) and its automatic conception in collaboration with the CIAN team.

1. MPI Application Task Integreation Platform

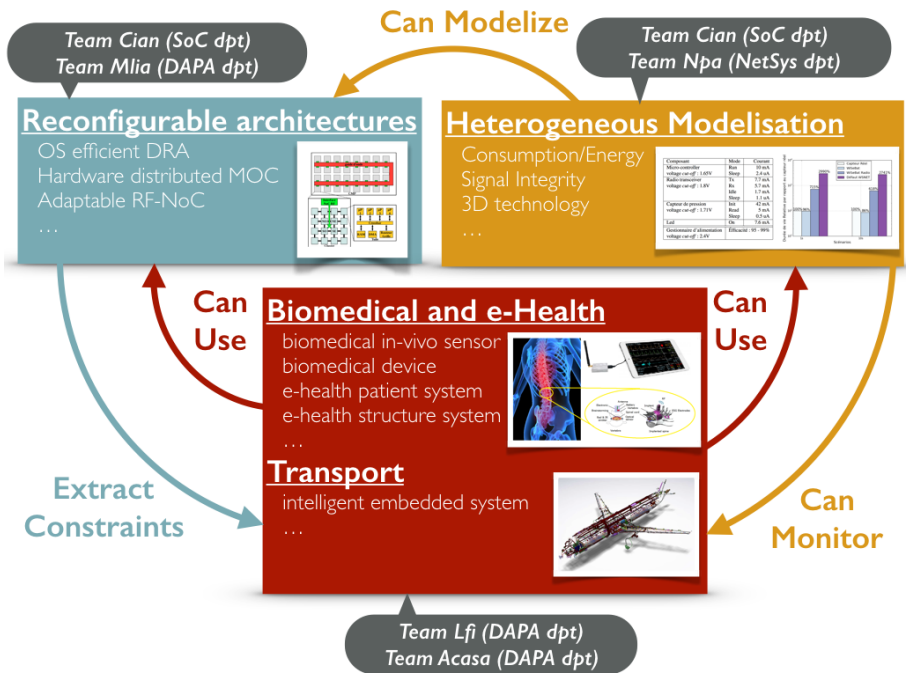


FIGURE 1.2 – SyeL thematics

Societal Challenge

Biomedical/e-Health We developed a strong activity in the biomedical and e-Heath domain, thanks to strong interactions with practitioners from APHP hospitals (also members of Sorbonne University) and around Europe. Our scientific approach here is to consider all the levels of electronic equipments, from the sensor to the system, to implement smart systems for an assisted diagnosis.

AT THE SENSOR LEVEL, we define and design sensors able of measuring the bio activity. Between 2012 and 2015, we studied in the PEPS-CNRS FibroSES project, a model by impedance measurement of the fibrosis on an implanted electrode, [**SYEL-C-11**]. Since 2010, we are designing a spinal cord (SC) implant to monitor both oxygen and electrical activities (SpinalCOM project) through the design and the modeling of a multimodal communicating device, [**SYEL-C-7**]. This approach will locally and specifically establish the functional state of the SC in real time, which will provide a breakthrough in the aortic surgery, as well as in the care and therapeutic trials for SC.

AT THE DISPOSITIVE LEVEL, we work on a new generation of wireless video capsule for endoscopy (WCE). We have defined a project, NAutILES (Novel Autonomous Intelligent Localisable Endoscopic System), [**SYEL-C-13**], to develop a new gastro-intestinal medical device : an intelligent ingestible wireless capsule featuring enhanced and integrated new multispectral photonic imaging and automated diagnosis capabilities, suitable for the detection of gastro-intestinal diseases, [**SYEL-J-3**]. We work to implement image processing algorithms and machine-learning algorithms inside a WCE to detect in-situ colorectal lesions [**SYEL-C-12**].

At this level, we also run an activity on the monitoring and analysis of sleep activity (AEP project) to design an automated polysomnograph system (PSG) that embeds intelligence capable of interpreting recorded PSG signals, [**SYEL-C-8**]. To do so, we use rules inspired by the international recommendations, defined by the American Academy of Sleep Medicine (AASM) and we use symbolic fusion to merge information through several levels of abstraction. We have designed a personal sleep staging system using evolutionary algorithms and symbolic fusion, [**SYEL-C-9**]. The AEP project also has for objective the building of a database of polysomnograph examination recordings of several patients. It has also as an objective the definition of a gold-standard for sleep scoring.

AT THE SYSTEM LEVEL, we design medical equipments for telemedicine services that integrate medical image compression algorithms for transmission. Such equipments could be X-Ray portable machines. We propose an implementation of a wavelet based certified medical image encoder from the CIRA company, which

outperforms the current state of the art JPEG2000 compression standard in terms of both PSNR and SSIM. The encoder is implemented in a SoC based design with a FPGA, [**SYEL-J-10**]. We have realized a state of the art implementation of a real time hardware DWT [**SYEL-J-14**]. We also improved the original image compression algorithm to increase its efficiency in terms of image compression ratio and real time processing, in order to allow medical video compression [**SYEL-C-14**]. We work on specific logarithmic coding [**SYEL-C-15**] to perform and accelerate the computation of a logarithmic Discrete Wavelet Transform and achieve real time processing by replacing multiplications by additions. We used our knowledge on medical image and video compression to conceive a tool to compress a multi-modal flow of neurological examinations in a FUI project : SmartEEG. This project is based on a collaboration between 2 laboratories (LIP6 and ETIS), APHP and 4 companies (CIRA, Partelec, 2CSI, Acasia). We have worked on the production of a compressed and synchronous flow of ExG signals, sound and video with a frame rate of 100 Fps [**SYEL-C-16**]. We also used this tool to perform the detection of epileptic seizure based on expected activity measurements and neural network classification [**SYEL-C-17**].

Transport We worked on the Corail project from the CORAC [2]. We proposed to introduce a new kind of monitoring based on embedded simulation. This simulator, implemented in SystemC, monitors data traffic generated by key avionic devices in order to detect suspicious behavior such as missing data, unexpected communication or simply incoherent data, [**SYEL-C-10**].

2. Produits de la recherche et activités de recherche

Bilan scientifique

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	0	2	3	5	2	3	15
Ouvrages				1			
Conferences	15	19	18	10	7	1	70
Logiciels					2		2
Brevets					2		2

Sélection des produits et des activités de recherche

The selected publications are listed in annex as required by the committee.

As can be seen, our publication activity has slowed down in the last two years. This is due to the few HDR and professors in the team and their huge investment in administrative duties. Patrick Garda is now Scientific Director at ministry of research and Bertrand Granado is Director of the EDITE doctoral school. This lack of HDR limits the number of new Phd students in the team. Hopefully, Andrea Pinna had obtained his HDR in June 2017, Sylvain Feruglio in on the way to obtain it in autumn 2017. The low number of professors still remains however.

Faits marquants

During the considered period, we can emphasize some notable events. We have strengthened our relationships with international teams :

— With Instituto di Elettronica e di Ingeneria dell'Informazione e delle Telecomunicazioni of CNR with Giovanna Morgavi who was invited in the team in 2013. The purpose of this collaboration is to define a technique for the detection of sleep events based on phase plot. This collaboration led to a software deposit [**SYEL-B-1**].

— With ULB at Brussels, Dragomir Milojevic visited for an one month stay in 2016 and Anh Von Doan visited in 2013 for a 3 month stay. The purpose of this collaboration is the definition of a method for digital architecture design in the context of biomedical applications.

2. COnseil pour la Recherche Aéronautique Civile

— With the creation of a network around the video-capsule with Anastasios Koulouazidis from the Royal Infirmary of Edinburgh, Dimitris Iakovidis from the University of Thessaly and Roberto Sarmiento and Gustavo Marreco from the University of Las Palmas de Gran Canarias. An article was published in 2014 with this network [**SYEL-C-13**] and a proposal was submitted at the H2020 call on H2020-ICT-2017.

We have also been strongly involved in technological transfer activities :

— Based on the patent [**SYEL-B-2**], we have participated to the Wisebatt startup creation, launched by one of our former Phd Student Wilfried Dron

— We are currently in partnership with two SME for the transfert of our technology. Bodycap SME supports the transfert of parts of NAutILES project, and Bioserenity SME the transfert of parts of AEP project.

We also developed two softwares :

— MATIP that allows implementation of hardware tasks using MPI

— ExASySo that extract symbol features for sleep stage detection

3. Analyse SWOT

For the next five years, we want to stabilize our scientific activities and keep the same organization. Our three majors thematics still remain.

Reconfigurable Architectures In the reconfigurable architectures theme, we will continue our studies on the RF-NoC started in the WiNoCoD project with two directions. The first one is the integration of such a RF-NoC in a 3D stacking many core to interconnect a third level of memory cache. The second, is to study the integration of a RF-NoC in a FPGA to perform the higher level of interconnect. These two subjects will be conducted in strong collaboration with ULB [3] and IMEC that will bring to us the 3D stacking knowledge.

In parallel, we will design a complete heterogeneous MATIP platform, that will combine a software part, conceived with RISC-V processors from the University of Berkeley, and a hardware part. Both of these processing elements will feature MPI and will have the capacity to communicate with the other part and also to create new tasks, either software or hardware. One of our objective is to increase the level of description of an application, by using methods such as High Level Synthesis and Desing Space Exploration.

Heterogeneous Modeling In the heterogeneous modeling theme, we will continue to study the effects of technological parameters on analog integrated chip performances. We will take in consideration features like power consumption, time and also in conjunction with RF-NoC studies the impact of 3D stacking. One of the goal is to integrate in a Design Space Exploration the technology node as a parameter. As in reconfigurable architectures, a collaboration with ULB and IMEC will take place here. As a working example, we will also consider the BQJ photo-sensor []. In conjunction with this activity, and the startup Wisebatt, we will investigate a high level modeling technique of batteries to adapt the behavior of a system, based on an OS management. Prediction of energetic autonomy for embedded devices will be at the heart of this modeling activity. Finally, synchronous systems will be analyzed in a collaboration with the CIAN team on oscillators.

Societal Challenge

Biomedical/e-Health For biomedical and e-Health societal challenges, based on heterogeneous modeling, we will implement biomedical dispositive models, using neural network methods. In conjunction with these models, we will investigate energy harvesting for a biomedical dispositive based on a biological reaction, like heat generation. Using these models, we will analyze the opportunity for us to conceive and implement low energy ASIC.

We will continue to conceive sensors for the measurement of bio activity with the objective to suit the needs of a neurophysiologist or a surgeon, a project start in 2017 across a collaboration between L2E lab and APHP. ANR project or H2020 will be considered here to create a strong collaboration between several european laboratories and hospitals.

3. Université Libre de Bruxelles

At the dispositive level, we will continue the NAutILES project. A H2020 project has been submitted on the multi-spectral aspect. We will also work on a new modality : the microbiot. For this particular aspect, we participate to a SIRIC [4] from Sorbonne Universités project that will be submitted in september 2017 at INC [5]. In this call, several hospitals are involved (La Pitié Salpétrière, Saint Antoine, Charles Foix, Tenon, Rothschild) and also laboratories from UPMC.

In parallel, a maturation project with the SATT LUTECH is ongoing to design our first prototype of a WCE. This work will be done with a SME named Bodycap. We will consider here also to find another partner in the network like EIT_Healht and EIT_Digital close to CMOS technology to enhance the WCE in terms of power processing, power consumption and real time.

Finally we will continue to develop our european network on WCE.

In the sleep monitoring and analysis (AEP) activity, we will formalize and make interpretation of medical knowledge in collaboration with the ACASA team. A maturation project is also ongoing with the SATT Lutech and the Bioserenity SME. We will also consider new collaborations and new partners, like Essilor and Air Liquide, in EIT_Health and EIT_Digital networks.

At the system level, our work on image compression will be reinforced with a collaboration with the LISITE laboratory. We will work on finding the best compression rate for biomedical applications, we will define specific embedded architectures to integrate image compression in nomad medical tools. Finally we will also work, in conjunction with the works in NAutILES and AEP, on automatic pathologies detection.

Transport In the continuation of the Corail project, where we designed a complete platform to perform monitoring based on embedded simulation, we will investigate use cases and stress our proposition to validate it. We will also investigate the impact of multi-core processors in avionic systems by taking into account the operating system.

4. SItes de Recherche Intégrée sur le Cancer
5. Institut National du Cancer

DEPARTEMENT SYSCOMP

1. Présentation du département

Within experimental sciences, computer science has a very particular characteristic : it generates itself some of its most central objects of study, such as the internet or large software. De facto, computer scientists know perfectly well the atoms and cells composing these objects : they actually design, document, build, and... assemble them. For the internet, the atoms are routers, cables, protocols. For large software, the cells are programming languages, function libraries, compilers. Computer scientists may easily zoom in and zoom out, from the electrical signal to the global system, they may refer to documentation, maps defining the architecture of the object or even source codes. These key computer science objects are indeed engineered objects, which makes computer science a priori very different from natural or social sciences.

Yet, these computer science objects are far from being mastered ; this is particularly clear for large software, whose bugs pace our interactions with the digital world and can cause disasters. More generally, they raise challenging questions regarding their structure, behavior, resilience, performance... which one is unable to answer only from the detailed knowledge of their construction. Indeed, this construction, far from strictly following the vision of a single and consistent architect, actually is rooted in long time periods (decades for the internet and large software), and most decisions regarding them are not centralized (they are distributed). From these processes emerge systems consisting of millions or billions of building blocks a priori simple, but whose collective behavior escapes our understanding.

In this situation, the naturally bottom-up vision (from the elementary bricks to the global system) of the computer scientist is no longer enough. It must be supplemented by a top-down approach, based on measurements, probes, in silico or IRL experiments, observation. The history of computer science has experienced several revolutions stemming from such approaches. One can think of the emergence of RISC processors in the 1980s, for example, following an analysis of the set of processor instructions used in practice ; or measures of the internet since the 1990s that are constantly contradicting the intuition of its designers. But top-down approaches remain largely under-considered in computer science, seen as an abdication of our position as Creators, very little present in the courses : the emphasis is on design of systems, not on analysis of existing ones.

The Complex Systems department at LIP6 proposes to tackle this challenges by putting together both top-down and bottom-up approaches in the design and study of computer science objects. Its research is conducted by 35 permanent researchers grouped into three teams : APR (methodological aspects based on algorithmics and programming theory), ComplexNetworks (focused on structural aspects with key formalism graphs and networks) and MoVe (study, design, security and verification of complex software).

2. Politique scientifique

We created the Complex Systems department from the RSR department by the end of 2013, and from a common vision of specific challenges that we decided to tackle together, as explained above. Since daily scientific activity and strategy is leaded by teams, the role of the newly created department was to promote this scientific vision among its members, and to create bounds with other LIP6 and UPMC structures as well as national and international ones.

The mere definition of our scientific area was at the core of tremendous discussions before its creation and at its beginnings. Indeed, the terms *complex systems* may refer to many different objects in physics, mathematics, and even in computer science. We needed a precise consensus on our common goal, and the convergence of bottom-up with top-down approaches detailed above finally was key to the emergence of the department itself. It

emphasizes the need of both practical and fundamental approaches and progress, in a combination of theoretical studies and empirical ones that are a hallmark of our department.

Complementary to the team-directed work described in next pages, we therefore developed several joint projects, some of which with visible impact : we submitted joint ANR projects and LIP6 projects, hired several interns co-advised by members of different teams, and published papers co-signed by members of different teams. We also had discussions on the department management and defined clear priorities for temporary and permanent reserchers hirings, in full agreement of all teams and with a consistent vision over time. Each team has its own seminar, open to all members of the department (and there are indeed attendees of other teams). In order to avoid increasing the number of meetings, we plan to make a joint issue of these seminars every two months, and to have a thematic workshop of one day every year.

In addition to this internal and local structuration of the department, we put much emphasis on its insertion into the national and international contexts. For instance, we are members of the board of directors of the ISC-PIF (Institut des Systèmes Complexes de Paris - Ile-de-France), RNSC (Réseau National des Systèmes Complexes), IRILL (Initiative de Recherche et Innovation sur le Logiciel Libre) and CSS (Complex Systems Society). We are among the main contributors of the *Feuille de route des systèmes complexes*, a collaborative document that defines priorities for complex systems research over years. Last but not least, we are involved in the boards of editors of major journals the field, like Journal of Complex Networks, Advances in Complex Systems, or Social Networks Analysis and Mining. We also promote complex systems in more generalist journals, in recruitement committees (notably at CNU and CNRS), and we organized some of the main events of the area, like for instance AofA 2014 (Analysis of Algorithms), ASONAM 2015 (IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining), POPL 2017 (Principles of Programming Languages), and the yearly Model Checking Contest.

Thanks to these efforts, in only 4 years the department became a key actor of the complex systems area in France and abroad, with an important national role and an international visibility. Its internal management relies on concertation and consensus, with collective decision-making processes and a deep involvement in LIP6, CNRS and UPMC structures.

EQUIPE APR

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The APR team was created 9 years ago by Michèle Soria, who was at its head until June 2016. It is composed of 5 professors (including 2 Emeritus), 6 assistant professors (including 1 detachment at ANSSI), 2 permanent CNRS researchers (1 DR +1 CR). Concerning the non permanent researchers, there are 2 post-docs and 10 PhD students. The APR team has a scientific unity around Algorithmics and Programming Languages. People with different backgrounds gathered in 2008 for historical reasons, and since then, many interactions have been developed at different levels, to import algorithmic and combinatorial methods for studying languages, to adapt algorithms to languages, to use complex algorithms for testing parallel and distributed programs. The team members are working and publishing together, including the new colleagues recruited since 2010.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	9	9	9 (+1 emeritus)	9 (+1 emeritus)	8 (+2 emeritus)	8 (+2 emeritus)
Researchers	1	1	2	2	2	2
PhD students	8	5	5	6	9	11
Others	2		2		3	3
Total	20	15	19	18	24	26

Grants	2012	2013	2014	2015	2016	2017
International				23 Ke	277 Ke	408 Ke
National	100 Ke	92 Ke	57 Ke	25 Ke	34 Ke	34 Ke
Industry	45 Ke		16 Ke	58 Ke	50 Ke	80 Ke
Others						
Total	145 Ke	92 Ke	73 Ke	106 Ke	361 Ke	522 Ke

Politique scientifique

The APR team works on thematics dealing with algorithms and programming languages, from both theoretical and practical aspects, combining the language approach and the algorithmic requirements in a wide range of applications, such as programing new machine architectures, developing languages for concurrence and mobility, software testing, static analysis. More generally, in the APR team, we apply the common trend of the SYSCOMP department of linking top-down and bottom-up approaches to languages and algorithms by developing formal and mathematically-founded methodological approaches, in the domains of semantics and combinatorics, in order to improve the quality and reliability of software and achieve concrete goals such as development of software prototypes and industrial applications.

2. Produits de la recherche et activités de recherche

Bilan scientifique

We follow a dual approach to achieving safe and efficient applications by combining programming languages (design, implementation and program analysis tools) and algorithms (modeling, complexity analysis,

random generation) through theory (formal approach in semantics and combinatorics) and practice (concrete objectives producing prototyping softwares and industrial applications).

Our domains of applications cover the following themes : Web development, Critical softwares, Concurrency and mobility, New hardware architectures, Quantitative study of actual data. We have developed strong links with IRILL (Initiative for Research and Innovation on Free Software) and are deeply involved in teaching, from early undergraduate courses to specialized Masters (STL, MPRI, ISUP) focusing on education for and through research.

The APR team is based on three axes following a common research theme around concurrency :

1. Combinatorics and Analysis of algorithms
2. Languages and Tools for Critical Softwares and Web Applications
3. Algorithms and Languages for Multicore Processors

Combinatorics and Analysis of Algorithms *Description* : In the APR team, we develop combinatorial methods for analysing the performances of algorithms and randomly sampling complex structures. Our methods allow to quantify the complexity of algorithms and efficiently generate combinatorial objects. Our research both aims at extending the expressiveness of the models, and adapting the methods to real application domains : information retrieval in massive data, robustness tests or software validation, modeling of graphs, quantitative study of concurrency. We lead several ambitious projects such as non-uniform sampling, sampling with multi-parameter specifications, graph computing with GPU programming as well as quantitative analysis of programming languages.

Positioning in the local / national / international context : The French community working on discrete random structures and the analysis of algorithms is distributed in several teams federated in the GDR-IM Group ALEA. Moreover, the ANR MAGNUM was an important cement for the Parisian teams. At the international level, our activity is well recognized, as evidenced by our participation in several program committees of international journals and conferences. We have also created an international GDR ALEA-Network, sponsored by CNRS, bringing together several European universities (Wien, Stockholm, Oxford, Munich) by organising meetings, exchanges of young researchers ...

Languages and tools for Critical Softwares and Web Applications *Description* : The APR team works also on design and implementation of programming languages from the point of view of expressiveness, reliability and efficiency. We aim at building a complete chain from computation models (processes calculus, type systems, static analysis, ...) to actual implementations. Theses works focus on two different application domains. On the one hand we are interested in the development of tools for the certification of critical systems (tests and static analysis by abstract interpretation). On the other hand, web-based technologies offers an excellent laboratory for new ideas, such as computational migration and synchronous-reactive programming for multi-tier applications. Both domains revolve around component-based systems which strongly benefit from a top-down vision (global control, type system, monitors, multi-tier design) in a way reminiscent of SYSCOM point of view.

Positioning in the local / national / international context : The part about programming in the diffuse Web was concretized by an academic project PWD ANR and then extended by an other project UCF FUI with industrial partners. A new academic project has started this year on analysis for a safe and efficient Web programming (APRES/PEPS). At the level of critical applications, two ANRs revolve around component certification (CERCLES) and the combination of abstract interpretation and constraint programming for floating calculation (COVERIF). Several bi-lateral collaborations (Imperial College - UK, Academica Sinica - Taiwan) spread the influence of the team, especially through the ERC Consolidator project MOPSA of Open and Modular platform for the static analysis for dynamic languages and web applications that has just started.

Algorithms and Languages for Multicore Processors *Description* : A real need for abstraction arose out of the democratization of parallelism and concurrency. These abstractions should facilitate parallel composition while allowing a compiler to generate efficient code for a given target architecture. These abstractions must also be clearly specified and allow the static analysis of a number of properties to ensure the safe execution of the programs created. The great strength of our team on this research theme is the complementarity of

the work we have carried out in recent years. On the one hand, we have a recognized expertise concerning advanced algorithmics and parallel programming techniques, particularly in the shared memory model and data-parallelism. This concerns, on the one hand, the abstraction of the hardware, CPU and GPU, and the algorithms adaptation for these specific hardware and on the other hand the tools of static analysis by abstract interpretation for concurrent programs for shared memory models as well as for communicating processes. In a complementary way, we have studied concurrency and mobility based on π-calculus.

Positioning in the local / national / international context : In the GDR GPL, the team participates in the Lamha group for high-level parallelism and the LTP group for static analysis as well as the compilation group. It has several projects in parallel, both academe by the PEPs ALPACA and Graph GPU, and on the other hand with academic / industrial collaborative projects OPENGPU / FUI and LCHIP / FUI. At the international level, this activity is carried out via bilateral (Brazil) and larger European projects such as ASSUME (39 partners from 5 countries) on using multicore for critical embedded applications.

Common research theme : concurrency The main interactions of the team relate on concurrency (theory and practice), the following subjects may be particularly noted and lead to projects :

- Analytic combinatorics for the quantitative study of concurrent systems (A. Genitrini, F. Peschanski, M. Soria, M. Dien), project ALPACA (PEPS)
- static analysis by abstract interpretation of dynamic languages and concurrent programs (A. Miné, T. Suzanne, V. Botbol), projects ASSUME (ITEA), MOPSA (ERC)
- interactions of Web applications using types sessions (R. Demangeon), project APRES (PEPS), and languages for multi-tiers programming (R. El Sibaie, Ch. Queinnec, E. Chailloux), project UCF (FUI)
- parallel programming and algorithms (B-M. Bui-Xuan, M. Bourgoin, E. Chailloux), project GRAPH-GPU (PEPS)

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	10	4	3	11	9	4	41
Ouvrages/chapitres			1		1		2
Conférences	15	8	11	12	12	8	66
Logiciels	3	1	1	2	1		8
Brevets							

All our softwares production is distributed under open source licences, so we have no patent. We have counted softwares from their first apparition, but except for two of them, they continue to evolve and are maintained.

Sélection des produits et des activités de recherche

We mainly select the productions around concurrency as previously described in section "scientific description" on : Analytic combinatorics, Static analysis, Web programming, Algorithms and parallel programming (with the following productions : articles [**APR-J-1**, **APR-J-2**, **APR-J-3**, **APR-J-4**, **APR-J-5**], proceedings [**APR-C-1**, **APR-C-5**, **APR-C-2**, **APR-C-3**, **APR-C-4**] softwares and projects which are completely described in appendice 4. We add a few other publication on using constraint programming for static analysis [**APR-C-6**], on memory for safe serialization [**APR-C-13**] and resources analysis [**APR-C-10**] and resolution of polynomial systems [**APR-C-14**].

Faits marquants

We are involved in many international collaborations, leading to joint papers, leadership of international projects (ASSUME/ITEA3, MOPSA/ERC Consolidator Grant for A. Miné), participations in programs committees, organizations of colloquia (some in Jussieu : AOFA 2014, POPL 2017). We have strong links with international groups working on similar subjects, which results in many exchanges of permanent and non permanent researchers. We have bi-lateral collaborations with Wien (PHC Amadeus project, A. Genitrini), Rio de Janeiro (Sorbonne Universités project, E. Chailloux), Taiwan (ANR-MOST project, A. Genitrini), or directly

with visiting colleagues : R. Demangeon with Pr. Y. Nobuko (Imperial College), A. Valibouze with Pr. Y. Ihsen (University of Sfax) and with Pr. A. Basiri and S. Rajmani (University of Damghan - Iran), ...

We (M. Soria) created the CNRS GDRI "Alea-Network" (2015-2019), an international scientific coordination network, gathering researchers interested by discrete random structures from Wien, Oxford, Munich, Stockholm, with an objective of working together, structure the European community and facilitate its integration in European projects.

We participate on many national academics projects : MAGNUM/ANR (lead : M. Soria), ALPACA/PEPS (F. Peschanski/A. Genitrini), PWD/ANR (C. Queinnec), GRAPHGPU/PEPS (lead : B-M. Bui-Xuan), APRES/-PEPS (R. Demangeon), COVERIF ANR (A. Miné), working groups as GDR IM and GDR GPL (M. Bourgoin won the thesis prize on 2015) and integrating Institut Univeristaire de France (A. Miné). We have a strong involvement in CNRS (J. Mairesse is deputy director at the INS2I since 2014). We have also important collaborations to the industrial world with collaborative projects OPENGPU/FUI, CERCLES2/ANR, UCF/FUI, LCHIP/FU (E. Chailloux), organizations of colloquia (OSIS 2016, OSIS 2017), participation to the SYSTEMATIC cluster (mainly inside the thematic group on "Logiciels Libres" Free Software), to produce new innovative softwares distributed in Open Source.

Since December 2015, the "Initiative pour la Recherche et l'Innovation sur le Logiciel Libre" (IRILL - the Center for Research and Innovation on Free Software) is localized on Jussieu campus. We participate to all its activities : seminar organization, projects and direction (deputy director : E. Chailloux).

There are strong links between our research and teaching activities. For example we are in charge (R. Demangeon) of the first undergraduate computer science course with 900 students and we created a MOOC on "recursive programming" (C. Queinnec). We also developed new courses and an international program on bachelor degree, we are in charge of the master-STL-Software Science and technology (140 students), of the computer science studies at ISUP (Institute of Statistics, 180 students), and of the Research master MPRI.

3. Analyse SWOT

Strengths. Strong cohesion of the APR team, in the dynamics of its creation, enhanced by excellent recruitment. Multiplicity of skills for addressing a diversity of original research directions. Joint work within the team. Fundamental research accompanied by tools development in Open Source.
Strengthening of the "languages" component with A. Miné's coming (contracts, docs, post-docs)

- Synergies with several teams in LIP6. Numerous national and international collaborations with academic teams. Animation of scientific community at national level (deputy director of CNRS-INS2I, Research and Innovation on Free Software at IRILL, ALEA GDR-IM, CNU) and international level (GDRI, ERC, AstréeA). Leadership in Abstract Interpretation (Antoine Miné : ERC and IUF).
- Strong implication into teaching activities at Bachelor and Master degrees in UPMC, main responsibilities : first bachelor course (900 students), master STL-Software Science and Technology (140 students), CS studies at ISUP (180 students). Important involvement in the Research Master MPRI.
- A dozen of former research students have interesting careers in academic research (Paris-Est, Orléans, Grenoble, Nice, Princeton) ; in GAFA companies (Google, Amazon, Facebook) ; Startups (Cryptosens, BeSports, OCaml pro) ; other companies (Airbus, Esterel-Technologies/ANSYS, Kyriba). One of our colleagues, P. Trébuchet is now in detachment at ANSSI (French Network and Information Security Agency ANSSI).
- A lot of old and sustainable industrial cooperations, with collaborative research contracts, CIFRE grants, participation in *Systematic*, the working group on Open Source. An important software production in Open Source, strong links with IRILL (co-direction).

Weaknesses. The team diversity, which is its main strength, can also be a weakness. We aim at maintaining our effort toward scientific cohesion and integration of our research themes.

- The balance that we have maintained between the different thematics and between permanent and non-permanent researchers is threatened by future departures : several MdC are about to defend their HdR (4 HdR planned by 2018), and 2 professors recently became Emeritus.
- Finally, several members of our team have been carrying, for years, heavy administrative tasks in LMD and Research ; departures force us to reorganise.

Opportunities. The following are opportunities to create new collaborations and disseminate our work more widely :

- diversity of domains of expertise and research axes ;
- collaboration with numerous academic teams and industrial actors ;
- very good integration in the industrial environment via the Systematic cluster (thematic group on open source software) ;
- numerous connections at national and international levels.

Threats.

- Because Web development and multi-core computing rely on always evolving technologies, the risk of having the applications of a research project becoming out of fashion is real and requires adaptability and continuous awareness of recent industrial developments.

4. Projet scientifique à cinq ans

Our main objective is the application of our methods to analyses of more complex realistic worlds : programming and verification of realistic programs, with mainstream languages and current hardwares ; analysis and random generation of complex structures in the context of large realistic data. A few projects are already planned to follow these trends.

- on mainstream languages and domains :
  - MOPSA (ERC Advanced Grant, 2016-2020) for “Modular Open Platform for Static Analysis”. The Mopsa project aim to bring static analysis to the next level : target larger, more complex and heterogeneous software, and make it usable by engineers to improve general-purpose software. We focus on analyzing open-source software which are readily available, complex, widespread. The scientific challenges we must overcome include designing scalable analyses producing relevant information, supporting novel popular languages (such as Python), analyzing properties more adapted to the continuous development of software common in open-source.
  - COVERIF (ANR project, 2015-2019) for “Combining abstract interpretation and constraint programming to verify critical properties of embedded programs with floating point computations ”. The aim of this project is to address this challenge by exploring new methods based on a tight collaboration between abstract interpretation (IA) and constraint programming (CP).
  - APRES (CNRS PEPS, 2017) for “Analyses pour une Programmation Réticulaire Efficace et Sûre “. APRES project proposes to apply formal methods to check concurrency properties to the modern Web programming. The kind of properties to check are structural, i.e. dependant of the interaction schemes of the web application and are concerning on the quality of service.
- on current hardwares :
  - ASSUME (2015-2018) for “Affordable Safe And Secure Mobility Evolution”.
    Currently the single most important roadblock for this market is the ability to come up with an affordable, safe multi-core development methodology that allows industry to deliver trustworthy new functions at competitive prices. Assume provides a seamless engineering methodology to overcome this roadblock. The problem is addressed on the constructive and on the analytic side. For efficient construction and synthesis of embedded systems, the project provides new tools, standards and methodologies to cover most of the challenges by design.
  - LCHIP (FUI 2016-2020) for “Double-Core SIL4 architecture aimed at low-cost safety critical automation” LCHIP project aims at largely faciliting the development of safe applications to high level of criticality. Our part is to validate a new execution model on micro-controllers by following a virtual machine approach, including concurrency between two micro-controllers which run two different instances of the same specification.
  - GraphGPU (CNRS PEPS, 2016-2017) project proposes to use scientific multi-thread GPU programming to speed up the runtime of implementations of parameterized graph algorithms based on cliquewidth related parameters such as rankwidth and booleanwidth.

Moreover new projects, on concurrent and parallel algorithms/programs/languages, will emerge from new synergies within the team, but also with other LIP6 teams (CN, ALSOC, RO, WHISPER, ...), french laboratories (IRIF, DI, CEDRIC, LIPN, LIGM, LIX, Verimag, LIFO, ...), INRIA (Paris and Sophia), and international

collaboration (mainly with Austria, Sweden, UK, Taiwan, Brazil and international GDR ALEA). Our investisment in computer science education at UPMC facilitates connections with a lot of small and big companies, where students from master (STL, Software Science and technology) but also Phd students from our team find good professional opportunities. In conjunction with IRILL, all of our software developments will be distributed on open-source for a widespread diffusion and to help their adoption in open-source communities.

EQUIPE Complex Networks

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

The team was created in 2008. In 2012 there were five permanent team members. Three left due to promotions, and we hired one *maître de conférences* in 2014, leading to a current number of permanent team members of only three.

Tableau des effectifs et moyens de l'équipe

Workforces	2012	2013	2014	2015	2016	2017
Professors	3	2	2	2	1	1
Researchers	2	2	2	2	2	2
PhD students	7	6	7	7	8	7
Others	1	1	3	3	3	3
Total	13	11	14	14	14	13

Grants	2012	2013	2014	2015	2016	2017
International	118 kE	90 kE	0	0	0	19 kE
National	158 kE	91 kE	125 kE	101 kE	170 kE	105 kE
Industry	0	16 kE	85 kE	79 kE	75 kE	0
Others						
Total	275 kE	197 kE	210 kE	180 kE	246 kE	124 kE

Politique scientifique

Complex Networks occur in many contexts : the internet, the Web, social, biological or judicial networks are examples of these objects, which can be modelled as graphs. Our team studies transverse questions about these objects, and in particular their measurement (aquirement of information about these graphs), metrology (induced bias on the structure by the measurement method), analysis (statistical or structural description), modelling (generating synthetic graphs sharing observed properties), as well as algorithmic questions (either new questions arising from the context or the taking into account of the very large size of these graphs).

Our approach consists in going back and forth between fundamental questions and applied problems, transverse fundamental questions arising from the study of applied problems, and the development of methods to answer these questions leading to applied results, as well as a validation for the relevance of the proposed methods.

We are close to academic teams dedicated to graphs, algorithms, networks and data, as well as teams in other academic fields, such as sociology or law, interested in particular instances of complex networks. We are one of the important teams on this topic, and our objective is to place ourselves in a leader position, in particular *via* excellent results and publications. We are also collaborating with companies interested by our results, both through common grant participations (such as ANR grants) and direct collaborations.

2. Produits de la recherche et activités de recherche

Bilan scientifique

Our team's research belongs to several overlapping family of questions, spanning a wide spectrum from fundamental to more applied questions.

An important part of our work has been devoted to the study of **dynamics** on/of networks. Two types of dynamics may occur : the network structure itself may evolve with time, with the appearance and disappearance of nodes and links ; or dynamic phenomena, such as rumour or disease propagation, may also occur on networks. In this last case, the dynamics affects the nodes' states.

We have introduced the *link stream* formalism, that allows the description of objects which are not naturally modelled by evolving graphs, but consist instead of temporary interactions. For instance, in email exchanges, each email is an interaction between the sender and the recipient, occurring at the time at which the email is sent. This formal framework allows to extend the notions existing in graph theory to these objects, leading to a natural and powerful descriptive vocabulary. We have also studied several new questions on link streams, such as node importance or the question of reachability. Our work was grounded on several practical cases, including email exchanges, physical proximity between individuals, IP exchanges, and have in all these cases reached a better understanding of the structure of the objects under study.

We have been interested in two related questions : link prediction (predicting which links will occur), and recommendation (propose products to customers based on the purchase history). We have introduced structural as well as time-based metrics that are relevant for these problems. We have also designed a machine learning method in order to combine these various kinds of information to predict links in the specific context of large graphs. We have shown that this method can improve the state of the art in a wide range of contexts.

Concerning propagation phenomena, the usual approach consists in modeling these phenomena with simple rules, stating how a node's state may influence its neighbours' states. We have used an orthogonal approach and studied real-world propagation phenomena, in order to describe them. We have shown in particular that one classical model used in this context, namely SIR, is unable to account for most of the characteristics of real spreading cascades, even when considering natural extension accounting for heterogeneous behaviours. This negative result has raised an alert against the widespread use of this model in the community. However, we also showed that integrating temporal information in an heterogeneous version of SIR improves its relevance.

Acquiring data on propagation phenomena is a difficult task because both the propagation itself and the network it happens on have to be acquired. We have managed to do so in four main cases : peer-to-peer (P2P) networks, phone calls and blog networks, and Twitter. We have shown that there exist two different time scales for these phenomena – the extrinsic scale corresponds to absolute time, and the intrinsic scale for which time is proportional to the number of structural changes in the network – and that the study of these two time scales allows to understand the interdependancy between propagation and structural dynamics.

Computing properties of interest on large (possibly dynamical) networks naturally raises strong **algorithmic questions**.

A *community* is a set of nodes that are densely interconnected but have few links to nodes outside of this group. Our team has had a strong activity on this topic since its creation, and we have during this period gone beyond the classical problem of graph partitioning. We have first considered the notion of *ego-centered* community, i.e. the community to which a given node belongs. Proposing a method to solve this problem led to the understanding that the community structure is much richer (and more complex) than what the state of the art thought. We have also proposed a relevant method for computing communities in bipartite networks. We have also been interested in communities in the dynamic case. In particular, we have proposed a method for detecting dense groups of links in link streams, which are groups of links involving few nodes, during a restricted time period (one may think of phone calls between colleagues working on a given project, for instance).

We have also proposed algorithms for the computation of two notions that we have adapted from classical graph theory to link streams : the question of reachability and computing maximal cliques.

Several of our works are concerened with **security** issues. We are involed in the question of **event detection**, where we consider as an event any observation that is statistically outside the norm for a given network. In the case of IP trafic between computers, such events may correspond to attacks. We have proposed methods for event detection in dynamic networks. In the case of IP trafic, we have moreover designed a *surgical removal*

of events method, that allows to identify and remove from the data the nodes and links involved in the event, which for instance allows to detect other events appearing at the same time.

The study of propagation phenomena is also related to security issues : in the case where a disease spreads in a population, an in-depth understanding of the underlying mechanisms helps in stopping or hindering it.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	3	7	8	4	5	1	28
Ouvrages							
Conferences	9	9	14	10	4	4	50
Logiciels	2		3	3	3		11
Brevets							

Sélection des produits et des activités de recherche

We present our selection of publications and other research production in Annex 4.

Faits marquants

Our team has a strong **interdisciplinary** activity, for two main reasons. First, the objects we are interested in arise in part from other fields (as well as from different sub-fields in computer science). Second, other scientific fields provide methods for the study of complex networks.

We have strong links with several fields in humanities. We have developed in particular a pluridisciplinary collaboration with researchers in law aiming at analysing corpus of judgements from the network science perspective. Indeed, judges motive their decisions by explicit references to former judgements and articles of law, which can be seen as complex networks. In this context, we have proposed new metrics allowing to highlight influential judgements and we have validated the outcomes by providing juridical interpretations of the results. We also have collaborated with sociologists within several common projects, past or ongoing. An ongoing collaborative project also involves geographers, with whom we will study questions about political geography and media sociology.

We have active collaborations with physicist on statistical methods for the description and the prediction of the dynamics of complex networks. We are also in close contact with several teams in mathematics trying to propose statistical models for these objects.

We are actively involved in the life on the community concerning these interdisciplinary aspects. We are a member of the newly created *Groupement de Recherches* GDR ARSHS ; we created a hub for the study of the modelling of temporal interactions at the *Institut des Sciences de la Communication* (ISCC), and one of our team members is part of the board of both ISCC and the *Institut des Systèmes Complexes* of the region of Paris, ISC-PIF.

Note that, besides interdisciplinar activities, complex networks occur as objects of studies for several fields of computer science. For this reason we are involved in collaborations with several other teams in LIP6, studying networking, systems, or formal aspects. In many cases these collaborations have been made official by an internal LIP6 project.

We have chosen to present as a highlight of our scientific contributions the results leading to the paper *Computing maximal cliques in link streams* [**CN-JM-1**].

A clique in a graph is a set of nodes that are all linked together. We extended this classical notion to the case of link streams : a clique is then a set of links *together with a time interval* such that all pairs of nodes are continuouly linked during this time interval. We presented a first algorithm for enumarating maximal cliques (i.e. cliques that are not included both temporally and structurally in any other) in link streams. Interestingly, though this algorithm is exponential with respect to the number of nodes, it is only polynomial with respect to the number of links, which may be much larger than n^2 and is often very large compared to the number of nodes.

We have shown that our algorithm is able to enumerate all maximal cliques in some cases of interest, in particular the link stream representing the interactions between students in a high school, captured by sensors.

We have shown that these cliques are a very interesting new tool for the better understanding of the structure of link streams of interest.

3. Analyse SWOT

Strengths. The Complex Networks field is an emerging field at the international level. We are one of the main team on this topic both in France and internationally. Our previous results place us in a very good position for proposing major contributions in the future.

Weaknesses. Our approach consists in tackling both real-world cases and fundamental approaches. This implies that we have a wide thematical spectrum, which is difficult for a small team.

Opportunities. Many topics concerning complex networks have very diversified contributions. The field lacks contributions that aim at unifying it, which is one of our goals. Moreover, our context offers many collaboration opportunities, both in different fields of computer science and in other academic fields.

Threats. Our topics receive a wide attention at the international level ; it is therefore difficult to stay competitive.

4. Projet scientifique à cinq ans

In direct line with the research we conducted in the previous year, our project combines **fundamental questions** that have emerged as particularly interesting during these last years, and important **societal concerns**. As explained above, these are not disjoint directions but, on the countrary, the societal questions will give rise to new fundamental contributions, which will in turn be validated by their applicability to real-world cases, while leading to concrete results in the cases we started from.

First, we are more convinced than ever that the field has a high need for contributions to the study of the **dynamics** of networks. Even though many contributions worldwide have addressed this topic in one form or another, there is no consensus on which methods and/or approaches are relevant for which cases, and the field lacks maturity. We believe we are in the right position to make important progress in this direction. In particular, the link stream formalism that we introduced is an important first step, and we intend to keep extending this formalism in the future, as well as use the properties we will introduce in this context for the study of real-world cases.

We have already started working on the problem of *link prediction* in dynamic networks. In the case of link streams, this question takes on a new meaning as links are ephemeral events and can occur several times, which is a problem that has not been considered by the literature until now. We want to explore a new approach consisting in taking into account both the structural properties of nodes, *as well as the history of the past links between them*, in order to propose efficient prediction methods.

Computing properties of interest on link streams naturally opens challenging **algorithmic questions**. Indeed, the extension of classical algorithmic notions and/or their computation raises new fundamental questions. For instance, while the notions of paths and distances have been adapted with success to the case of link stream, the notion of connectivity remains yet unsolved. Indeed, the property of transitivity is lost as there can exist temporal paths from a node a to a node b, and from b to another node c, while no temporal path exists from a to b. The field therefore requires both a relevant definition for connectivity and an efficient algorithm for computing the corresponding connected components. As another example, defining an extension of the notion of betweenness centrality is quite straightforward. However, its computation requires to compute the shortest paths going through a given vertex for a given time, *for all vertices and for all times*, which obviously cannot be done in a naive way.

Finally, the field is in dire lack of **models** for link streams, i.e. methods for generating artificial link streams sharing the properties of real-world link streams of interest. The classical approach when generating synthetic static networks consists in trying to sample at random with uniform probability a network among those which share a given set of properties. In the link stream case, straightforward random shuffling, both in the structural and temporal dimension have been proposed in the past. However, the andomized link streams exhibit properties which differ too much from the real streams to be usable in practice. We want to explore more constrained randomization processes that would yield link streams with relevant properties in regards to the real streams.

As already explained, the above questions will both fuel and be fueled by advances in some cases that are important **societal concerns**. First, we are involved in interdisciplinary projects [1] studying the **behaviour of users** in online systems and the influence of algorithms on this behaviour. These projects study filter bubbles and echo chambers, i.e. how recommender systems, for instance in web search engines or news site, which present different content to different users, tend to expose a user only to contents and/or users who already share their opinions, and possibly shift these opinions to more radical stances.

Second, our link stream analyses open pathways for progress in **security questions**. In particular, our work in event detection opens interesting directions for the detection of attacks in IP traffic, or for frauds in bank transactions. In particular, we plan to study how metrics that combine structural and temporal aspects can be used as input features in machine learning tools. Another approach is to develop pattern searching methods. The idea would be to be able to specify a behavioural pattern (describing a malevolent behaviour) by specifying interaction rules such as ; a must interact with b between an interaction with c and an interaction with d. We want then to design methods for searching for these patterns in a link stream, in the same way as the tool `grep` searches for patterns in text. We are involved in collaborations with two leading companies on this topic : Thales and Ercom.

1. The ANR grant Algodiv and the EU project ODYCCEUS

EQUIPE MOVE

DOSSIER D'AUTOÉVALUATION

1. Présentation de l'équipe

Introduction

In 2006, the team called "Systèmes répartis et Coopératifs" (*Distributed and Cooperative Systems*) split into MoVe (*Modeling and Verification*) and Regal (*Répartition et Gestion des Applications à Large échelle*). MoVe gathers staff members from several universities : Univ. P. & M. Curie (11), Univ. Paris Descartes (4), and Univ. Paris Ouest Nanterre (7). It also hosts 2 Emeritus professors.

MoVe contributes to two research areas. The first one concerns verification (mainly using model checking). The second one deals with new approaches for software development, considering modeling aspects and partially using formal methods. All our theoretical results are confronted with assessment of case studies. These past five years, the main applications domains we have investigated are verification of distributed algorithms, safety of autonomous vehicles, Web services, and ambient intelligence systems.

Some MoVe members have administrative responsibilities at UPMC (adjunct director of the "licence d'informatique", head of master "systèmes et applications répartis", one vice-president), at Paris Descartes (director of studies of the 2nd year of computer science at the IUT), or in Paris Ouest Nanterre (head of a master curriculum and one vice president).

Tableau des effectifs et moyens de l'équipe

Workforce	2012	2013	2014	2015	2016	2017
Professors (emeritus)	20 (2)	22 (2)	22 (2)	22 (2)	22 (2)	22 (2)
CNRS Researchers	0	0	0	0	0	0
PhD students	13	12	9	5	6	3
Others	1	1	4	3	2	3
Total (emeritus)	34 (2)	35 (2)	33 (2)	30 (2)	30 (2)	28 (2)

Grants (KE)	2012	2013	2014	2015	2016	2017
International	39.36	246.02	210.76	199.54	71.26	81.36
National	165.18	68.74	43.29	29.62	47.28	47.60
Industry	26.00	41.11	0.00	0.00	12.00	15.00
Others	0.00	0.00	0.00	0.00	0.00	0.00
Total	230.54	355.87	254.04	229.16	130.54	143.96

Politique scientifique

Ensuring reliability of complex systems is a difficult yet crucial task. In this domain, "classical" validation strategies (mostly based on testing) are not sufficient because of the interleaving between instructions executed on different threads or processes. Moreover, the increasing number of heterogeneous and concurrent entities, in the era of internet of things (IoT), makes this task even more difficult while it remains crucial since such systems are expected to handle mission critical, or life critical, functions.

In this context, our activity deals with both theoretical and practical aspects where new ideas are confronted with real cases studies. This forces us to implement new efficient algorithms to tackle combinatorial explosion. Thus, our tools regularly participate in competitions, or handle large benchmarks (for example, analysis of several versions of Eclipse to assess our contributions in software product lines).

2. Produits de la recherche et activités de recherche

Bilan scientifique

MoVe's activities deal with the modeling and analysis of behaviors in complex systems. We emphasize on quantitative or concurrent aspects of such systems. Investigated domains are :

- formal notations (as well as their complexity and decidability issues) to describe the behavior of concurrent timed or stochastic systems, which raises interesting semantic issues
- efficient algorithms for model-checking, capable to tackle the complexity generated either by instruction interleaving in concurrent systems, or by the manipulation of data (such as time),
- controller synthesis, used when a system cannot satisfy its specification, the underlying idea being the exploration of strategies able to restrict its behavior to meet the specification.

The techniques developed by MoVe are experimented in real situations thanks to case studies in various application domains : distributed algorithms, embedded systems (e.g. in cars or aeronautic systems), Attack and Protection Trees, protocols, ambient systems, etc. Some domains we also investigate raise interesting questions and thus, have been studied as a whole these past years :

- Process engineering : here verification targets two needs. The first one deals with the detection of a deviant behavior (compared to a given methodology). The second one concerns the identification of "repair operations" to come back in the main stream of the methodology. We have strongly participated in the elaboration of the OMG standard SPEM2 that defines a language suitable for process description, so, it is a basis to experiment our results.
- Software Product Lines : the goal is to refactor many variants of a program in order to reach a better architecture where the core functions, as well as the optional ones, are appropriately structured and identified.
- Mixing formal methods together with testing approaches in concurrent systems : the objective is to rely on formal methods to generate appropriate sequences of tests, without redundancy, that cover a set of pre-defined criteria. Experimentation is currently done on Web services, based on description of components (using languages such as BPMN, BPEL, or WSDL).
- Use of formal methods to validate ambient systems with an emphasis on planning and learning techniques.

MoVe's activity has resulted in the defense of 15 theses and 4 habilitations between 2012 and 2017. We also participate in the co-supervision of four Ph.D. students who are not members of LIP6 : H. Nguyen (Univ. Orsay, defended in 2013, co-supervised by P. Poizat and F. Zaïdi – LRI, Univ. Orsay), R. Khéfifi (Univ. Orsay, defended in 2014, co-supervised by P. Poizat and F. Saïs – LRI, Univ. Orsay), M. Benabdelhafid, (Univ. Constantine II, co-supervised by B. Bérard and M. Boufaida – Univ. Constantine II), Y. Duplouy (IRT SystemX, co-supervision started in 2015, co-supervised by B. Bérard and S. Haddad – LSV, ENS Cachan).

Below are some highlights concerning our scientific activity.

Capturing more semantics in a formal way. We propose a new subclass of hybrid automata extending Interrupt Timed Automata (ITA) [**MOVE-J-5**], formerly introduced to describe systems where tasks can be suspended and later resumed. Our new subclass, called Polynomial ITA [1] relies on hierarchically organized stopwatches (real-time variables evolving with a slope of 0 or 1 with time). It has a highly expressive power and features polynomial constraints and updates on stopwatches, while retaining good verification properties. In particular, the reachability of control states can be computed in 2-EXPTIME. This work resulted from a cooperation with PolSys, granted by the LIP6 project Parasol, and with two other laboratories (LACL and LSV).

Many verification tools can only handle a specific formalism, thus limiting the reusability of algorithms and data structures. We have proposed a general yet simple intermediate language to express discrete state semantics : Guarded Action Language [2] (GAL). This enables multi-formalism verification using model trans-

1. B. Bérard, S. Haddad, C. Picaronny, M. Safey El Din, and M. Sassolas. Polynomial Interrupt Timed Automata. In The 9th Workshop on Reachability Problems (RP'15), volume 9328 of Lecture Notes in Computer Science, pages 20–32, Warsaw, Poland, Sept. 2015. Springer

2. Y. Thierry-Mieg. Symbolic Model-Checking Using ITS-Tools. In Tools and Algorithms for the Construction and Analysis of Systems, volume 9035 of Lecture Notes in Computer Science, pages 231–237, London, United Kingdom, Apr. 2015. Springer Berlin Heidelberg

formation approaches from a variety of formalisms or domain specific languages. Tools operating at GAL level benefit from all the supported formalisms, and many simplifications and abstractions defined as GAL to GAL transformations (pre-processing for optimization) are also shared amongst the verification approaches.

Multi-technologies based model checking. These past years, we have elaborated new algorithms and technologies for model checking. Let us note three interesting results. The first one deals with the definition of a new class of Transition-based Generalized Testing Automata, a new form of automata inheriting from both Testing automata and Transition-based Generalized Büchi Automata ; associated explicit and symbolic model checking algorithms have proven their efficiency for stutter-invariant properties [**MOVE-C-9**]. The second one deals with the combination of two so-called "symbolic" techniques : those based on decision diagrams and those based on symmetry detection ; the proposed algorithms have proven their efficiency compared to other approaches when dealing with data-types [**MOVE-C-13**]. The last one concerns parallelization and optimization techniques for the evaluation of LTL properties ; techniques based on the decomposition of the formula into subcategories, as well as the use of lock-free data-structures led us to propose new efficient algorithms [**MOVE-J-3**].

We have developed new symbolic [**MOVE-C-9**] and hybrid [3] approaches for LTL model-checking using our powerful Hierarchical Set Decision Diagram engine. These tools leverage Spot from LRDE, to deal with formula translation and to drive the emptiness-check procedure. Our main solution engine for GAL is symbolic, and uses new equivalence classes aware operations to have efficient native symbolic decision procedures for arrays and arithmetic [**MOVE-C-13**]. We have developed more recently explicit, multi-core and partial-order reduction solutions for verifying GAL by establishing a bridge to LTSmin toolset of Twente, NL.

Boolean Satisfiability (SAT) and SAT solvers are nowadays widely used as backend engine for system verification (e.g. Bounded model checking). Hence, these last few years, we started investigating this domain from both exploitation of parallelism and symmetries aspects. Concerning the parallelism, we develop new approaches for SAT solving that benefit from modern many-core architectures. In this context we collaborate with the Regal team at LIP6 (expert in multi-core programming) and our first concrete result is a framework allowing to parallelize sequential SAT solvers [**MOVE-C-1**]. Concerning the integration of symmetries, we aim at capitalizing on our long expertise in the use symmetries for verification of complex systems to develop new approaches that outperforms previous ones. This work remains on progress since it started in October 2016.

Synthesis and quantitative verification. We have developed a collaboration with the NPA team at LIP6 on the verification and synthesis of algorithms for swarms of autonomous mobile robots (as instances of self-stabilizing problems). Existing distributed algorithms for such robots have been modeled as a model-checking problem and we discovered bugs [**MOVE-C-11**]. When robots evolve in a synchronous semantics, we have proposed an encoding of their behavior as a game, allowing to discover automatically distributed algorithms that are "correct-by-design".

Our work also concerns verification and control of timed systems. We extended the theory of Event-Clock automata, by defining a specific notion of zones adapted to this model, and developed widening operators to allow the forward analysis of these zones to terminate. In another direction, we studied controllability of timed systems when the specification is given in MITL, a syntactic restriction of the timed temporal logic MTL, and investigated under which restriction such a control problem can become decidable [**MOVE-C-6**] (collaborations with Université Libre de Bruxelle and Université d'Aix-Marseille).

We obtained several results on probabilistic opacity, a security related property aiming at measuring information leaks in the presence of partial observations and uncertainties [**MOVE-J-10**]. Probabilistic opacity was first studied for Markov chains in M. Sassolas's thesis, were several variants were computed. In the context of several collaborations (IST Austria, Ecole Polytechnique de Montreal, Femto laboratory in Besançon), his framework was extended to Markov decision Processes and Interval Markov chains, models where uncertainty is increased by adding non determinism to probabilistic transitions. We established the decidability status of various quantitative opacity problems for these models and, in the latter case, we also obtained monotonicity of the disclosure with respect to the refinement process [**MOVE-C-4**].

As a sequel of the Ph.D. thesis of Yann Zhang (defended in 2013), we have developed an activity about the safety of autonomous vehicles. This leads to the co-supervision of a Ph.D. student with LSV (ENS de Cachan)

3. A. Duret-Lutz, K. Klai, D. Poitrenaud, and Y. Thierry-Mieg. Self-Loop Aggregation Product - A New Hybrid Approach to On-the-Fly LTL Model Checking. In 9th International Symposium on Automated Technology for Verification and Analysis (ATVA'11), volume 6996 of Lecture Notes in Computer Science, pages 336–350, Taipei, Taiwan, Oct. 2011. Springer.

granted by IRT SystemX. The objective is to use statistical model checking to produce performance indices related to the probability of collisions for an autonomous vehicle controlled by a Simulink model [4].

Verification, test and composition of Web services. We develop a model-based approach to automated end-to-end functional testing of large scale heterogeneous services (seen as black boxes) and service architectures (seen as grey boxes). This approach is implemented as a service to be hosted on a cloud infrastructure. It relies on model-checking, temporal logic and probabilistic graphical reasoning to address optimization, efficiency and coverage of the produced tests [**MOVE-J-1**]. This work was carried out in collaboration with the Decision team at LIP6 within the context of the European FP7 MIDAS project.

Service composition enables the automated construction of value-added applications from an abstract composition requirements and reusable services. We have proposed new composition algorithms based on graph planning supporting on the one hand quality of service [5], and on the other hand the quality of information and its contextualization [**MOVE-C-16**].

Choreographies are a form of global, high-level specifications of a distributed web service composition. We have proposed an approach to generate Web services behavioral contracts from a choreography in a way that ensures the conformance of the distributed communicating service with respect to the choreographic specification [**MOVE-C-15**]. We support data by using symbolic models (in the sense of symbolic transitions), hence avoiding both over-approximation due to data removal and state space combinatorial explosion due to the size of data domains.

Business process engineering and verification. We focus on the modeling, execution and verification of software and business processes. BPMN is the de facto standard notation for business process modeling. We have developed the VerChor approach and tool for the verification of BPMN choreographies [**MOVE-J-8**]. This tool can be integrated into the Eclipse BPMN designer, thus allowing one to check for choreography realizability and conformance. We also have proposed evolution relations between business processes and the VBPMN tool allows us to check them on BPMN processes [6].

During the last five years, we have also tackled software process deviation detection and software process verification issues using first Alloy [7] then SMT approaches [**MOVE-C-10**, **MOVE-J-11**]. Our approaches and tools on software processes were validated in the context of the MeRGE ITEA Project.

Software Product Lines and code refactoring. We develop in our team an innovative extractive approach for defining and manipulating Software Product Lines (SPL). The objective is to migrate existing similar product variants created using the "clone-and-own" approach, into an SPL. So, instead of designing variability, we focus on variability inference. First, we consider static variability inference, based on structural informations of the product variants [**MOVE-C-2**, **MOVE-C-8**] (e.g. source code, models, requirements, documentation, etc.). Moreover, we also consider dynamic information inferred from execution traces. When such information is available, it is possible to deduce the behaviors of these related products, thus inferring the SPL ones. Combination of the two techniques [**MOVE-C-14**] is a main contribution in the domain that was developed in collaboration with the MLIA team at LIP6. These contributions are integrated in the BUT4Reuse tool.

In large software, detecting and removing unwanted dependencies is essential to refactor and reduce the cost of some changes by keeping them local. Unfortunately, even if more and more individual refactoring operations are automated in major IDEs, the breaking of unwanted dependencies is yet mainly performed manually. We have elaborated a solution to solve this problem. Proof of concept has been validated on a tool PUCK which is able to handle industrial-size software.

Verifying ambient systems. We collaborate with the SMA team about the implementation of an ambient intelligence guidance service to dynamically and concurrently handle the intentions of a given user [8]. The

4. B. Barbot, B. Bérard, Y. Duplouy, and S. Haddad. Statistical Model-Checking for Autonomous Vehicle Safety Validation. In SIA Simulation Numérique, Montigny-le-Bretonneux, France, Mar. 2017. Société des Ingénieurs de l'Automobile.

5. M. Chen, P. Poizat, and Y. Yan. Adaptive Composition and QoS Optimization of Conversational Services through Graph Planning Encoding. In Web Services Foundations, pages 423–449. Springer Verlag, 2014.

6. P. Poizat, G. Salaün, and A. Krishna. Checking Business Process Evolution. In 13th International Conference on Formal Aspects of Component Software (FACS), Besançon, France, Oct. 2016.

7. Y. Laurent, R. Bendraou, S. Baarir, and M.-P. Gervais. Alloy4SPV : A Formal Framework for Software Process Verification. In ECMFA 2014 - 10th European Conference on Modelling Foundations and Applications, volume 8569 of Lecture Notes in Computer Science, pages 83-100, York, United Kingdom, July 2014. Springer.

8. J.-M. Ilié and A.-C. Chaouche. Toward an Efficient Ambient Guidance for Transport Applications. In 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017), Leuven, Belgium, July 2017. Procedia Computer Science.

underlying contextual planning service is an innovative mechanism that manages possible actions under spatio-temporal contexts. We use formal methods, and more specifically algebraic notations, to specify scheduling information and describe various alternatives to reach a goal. The planning mechanism that is being developed is enforced with learning approaches deduced from past executions.

Données chiffrées

Production	2012	2013	2014	2015	2016	2017	Total
Journaux/revues	12	4	8	6	11	5	46
Conferences	15	25	26	23	16	14	119
Ouvrages et édition d'ouvrages	1	3	2	2	2	0	10
Chapitres de livres	0	6	1	0	1	0	8
Logiciels	2	3	6	5	6	5	27
Brevets	0	0	0	0	0	0	0

MoVe publishes and maintains the following software (listed in alphabetical order) : BenchKit [9], BUT4-Reuse [10], CosyVerif [11] (collaboration with LSV, LIPN and Université de Genève), ITS-Tools [12], Petri Net repository [13], PNML Framework [14], PNML2NUPN [15], PNXDD [16], Puck [17], VBPMN [18], and VerChor [19]. In the line "Logiciels", we count the major releases of these software that represent a global volume of about 650 KLOC. We also participate in the design of some of the Spot [20] algorithms (collaboration with LRDE) but not in their implementation issues. Prototype software (proof of concepts) which have not reached the distribution stage are not considered here.

Sélection des produits et des activités de recherche

Selected outputs and products from our research activity are presented in section I of the annex.

Faits marquants

In 2011, we started the first edition of the Model Checking Contest [21] that now reaches its 7^{th} edition. This international event is organized with colleagues from various institutions and gathers model checkers to compete on a common set of benchmarks enriched every year by the community (and now grouped into a separate database available to the community : Petri Net repository). Even if these benchmarks are strongly related to Petri nets and the ISO/IEC standardized exchange format (PNML), it attracted tools from other communities such as LTSMin (Univ. Twente). This event is more and more cited and the associated benchmark collection regularly used by colleagues to assess new algorithms and their implementation. Moreover, it has contributed to improve the quality of the model checkers which participate.

For several years, we have developed and enhanced our own kind of decision diagrams that were extremely competitive to capture the hierarchical structure of complex systems and allowed to gain orders of magnitude in the size of systems we could process. These past years, this structure has been used and adapted by several colleagues, in particular in Geneva and Toulouse. Tools developed on top of variants of these decision diagrams were confronted several times, for example during several editions of the model checking contest and were often ranked among the best in this contest.

9. https://lip6.fr/BenchKit.
10. https://lip6.fr/but4reuse.
11. http://cosyverif.org.
12. http://ddd.lip6.fr.
13. http://pnrepository.lip6.fr.
14. http://pnml.lip6.fr.
15. http://pnml.lip6.fr/pnml2nupn.
16. https://lip6.fr/pnxdd.
17. https://lip6.fr/puck.
18. https://lip6.fr/vbpmn.
19. https://lip6.fr/VerChor.
20. https://spot.lrde.epita.fr.
21. https://mcc.lip6.fr

We have also expanded our model checking libraries and tools with new techniques inspired and extended from existing work. As an example, we investigate the use of SAT and SMT solvers to verify systems (bounded model checking), integrated CEGAR (Counter Example Guided Abstraction Refinement) approaches, etc. This results in tools able to process several classes of algorithms in parallel, making use of multi-core capabilities in modern architectures. These works lead to the defense of two HDR : the one of Yann Thierry-Mieg (UPMC) and Alexandre Duret-Lutz (EPITA), a former Ph.D. student we strongly collaborate with.

Another important field investigated these past years is the optimization of SAT engines themselves. The idea is to exploit the intrinsic characteristics of satisfiability formulas issued from bounded model checking problems to to optimize their processing by SAT. In 2017, this new track was granted by a publication in the SAT conference [**MOVE-C-1**] and the submission of a tool for parallel SAT solving at the SAT competition. Other ideas are being investigated and aim at being combined in the next years.

These past years, we have experimented a radically new approach to deal with Software Product Line engineering. This is a "bottom-up vision" that, instead of designing variability, infer it from existing artifacts in software projects [**MOVE-C-2**]. It considers static variability inference for SPL extraction but also addresses dynamic variability inference from execution traces for a better precision. Experiments were handled with success on large bases of code such as Eclipse. This work lead to the defense of Tewfik Ziadi's HDR.

Finally, the topic of models and metamodels co-evolution has been successfully addressed in the past three years. This has been achieved in the context of automatically migrating OCL constraints after metamodel's evolution. This work was part of the FUI MONOGE Project and has led to the publication of 9 conference papers (5 of them are A ranked, one best paper awards) and two A+ ranked journals (IS and IEEE TSE). This work is also part of Reda Bendraou's HDR defended in Dec 2015.

3. Analyse SWOT

Strengths. MoVe is a well established team in its research area, gathering recognition from its group impact but also from individuals' one, given its record of well-cited publications in the best international venues. MoVe also has a long tradition of producing top and up to date software tools competing in performance and functionalities with the best ones in their areas. The team is very well integrated in the regional (MeFoSyLoMa) and national (GDR ALP, GPL and IM) research networks. As a large team, it also benefits from the breadth of its research topics coverage, both in knowledge and skills, as well as its members completing very well one another. Indeed, its mix of expertise in formal methods and software sciences is rather rare, especially in France.

Weaknesses. A recent trend, with the growing fierceness in the competition for research fundings, is both a drop in global research fundings and in the number of Ph.D. candidates. Another consequence is a slow down in the number of publications for members struggling to find funding and Ph.D. candidates with some under the threat to become officially non productive researchers from the CNRS standards. Finally, size has its drawbacks, the team being forced to find a balance between breadth and cohesion with project-based fundings regularly acting as centrifugal forces by orienting too tightly the research topics. Another point of attention is the geographical scattering of the team among three different universities which may impede cooperation among members (even if there are counter examples these past five years).

Opportunities. Modeling and verifying software and systems remain a top priority in computer science research as safety, correctness and reliability are still bitterly-sought qualities for an ever larger spectrum of critical software to business and society in general. This trend is not to be seen stopping as new areas emerge to challenge the existing results and tools (for example, ultra large-scale systems, IoT, cyber-physical systems, etc.). The high industrial demand for techniques and tools in the area provides MoVe with several opportunities to transfer and value its results (in software development processes and tools, refactoring, product lines, validation and verification, etc.). The software produced in the team can play a catalytic role to accelerate this and UPMC has largely strengthened its capability to help in transfer to industry. Finally, the current research evaluation process goes hand in hand with a similar evaluation and reform process for master programs, hence offering the possibility to integrate right away the research priorities of the team into master courses to attract new Ph.D. candidates on these topics in the next years.

Threats. Because of a lack in internal human resources devoted to the development of its software tools, MoVe remains too largely dependent upon its capability to find external fundings and hire engineers on its projects.

Also, given what seems to be an ever growing pressure for faculties to accept heavy administrative duties (undergraduate and graduate programs, executive positions within the university, national research networks management, etc.), several MoVe members have to devote a large share of their time outside pure research (and teaching) activities. Another threat is the "maîtres de conférences" of the team gaining experience, getting their Habilitation (HDR) and potentially leaving the team to get professor positions outside UPMC, which would result into a loss of knowledge and skills for the team since recruitment becomes more and more difficult.

4. Projet scientifique à cinq ans

Challenges. Our research foster on applications going from Cyber-Physical Systems (CPSs) to Business Processes (BPs) and address questions related to reliability and security.

The important advances in hardware and computer infrastructures offer particular growth potential of resources and computational power, pushing companies to deal with more and more complex applications. This results in highly distributed systems, with a very large size and many functionalities covering security and safety related concerns, and using different communication modes (synchronous vs. asynchronous) as well as heterogeneous modeling/programming languages. Such systems are referred to as Cyber-Physical Systems (CPSs) which are ubiquitous systems, applied in diverse domains including aerospace, energy, and healthcare and integrate software elements with physical entities such as sensors, actuators, that stand, for example in the so-called Internet of Things (IoT) or in automated cars. CPS management must thus deal with these physical entities and their behaviors, leading to many issues for the modeling, the verification, control and test of reliability and security properties. Business Processes and Web services, implemented as distributed systems, present specific communication features, including an evolving communication topology, hence a variable number of processes, which requires a parameterized approach.

The security problems addressed by our team are mostly concerned with information flow detection, where data confidentiality may be threatened by external agents, possibly interacting with the systems via partial observation. Another issue is related to automated deployment and self-maintenance of auto-adaptive applications on hybrid clouds, which also require data protection due to the mixing of public and private services.

Our objectives can be described along two aspects presented below. They concern the development of new models/formalisms taking into account some specific features on the one hand, and new algorithms for their analysis, control, monitoring and test on the other hand.

Models for behavior and variability. The systems we study are both *behavior-driven* and *variability-driven*. Variability is the ability of a software system or artifact to be changed, customized or configured for use in a particular context. Any variation concerning the software and physical entities affects the dynamics of the systems, which can lead to completely new behaviors.

A first research direction of the team consists in developing models of hybrid systems adapted to some particular applications, possibly including probabilistic aspects reflecting physical uncertainties. For Petri net based formalisms, we will participate in the revision of the PNML standard (ISO/IEC 15909) for Petri nets starting in 2018, where new extensions will deal with modularity and other additional features such as time and priorities.

Software processes aim at improving the quality and productivity of software development by encoding sets of well-known practices for implementing them using process models. For the composition of BPs, we intend to propose a generic model for parameterized workflow processes, and extend it to collaborations and choreographies. We also plan to develop model transformations applying to the BPMN ISO standard, with the aim to make the systems amenable to verification with SMT solvers. Stochastic automata might also be used in a project addressing separation of concerns in epidemiological modeling.

Since the considered systems are variability-rich systems, another track is devoted to the design of models able to tackle this aspect, using an approach based on extractive Product Line (PL). Such models should be able to integrate new elements (physical entities or artifacts) by successive refinements of an already extracted PL.

Our modeling activities also address the aspects related to the systems life cycle. This includes development processes of such systems as well as their refactoring activities. For development processes, our goal is to study the verification of these processes regarding the different aspects of systems such as control flow, data flow and resource management. Concerning refactoring, in addition to detect and remove unwanted dependencies our goal is to automate dependency breaking, generate coupling constraints and refactor product lines.

Finally, security models will be studied, both from a theoretical perspective with semantical issues for opacity on Markov Decision Processes and to deal with data protection in the context of auto-adaptive applications on hybrid clouds or information flow in operating system kernels. An investigation of models for Digital Right Management adapted to electronic documents is also planned.

Control, monitoring and testing. To achieve safety, or more elaborate properties like auto-regulation, is a notably challenging task, that will be addressed in several ways.

A core activity of the team concerns the fight against combinatorial explosion inherent to verification by model checking. New research directions were recently started, on the integration of SAT solvers, the parallelization of model checking, to be combined with partial order reductions, and the more prospective issue of adaptive model checking. These works will go on, using the models and properties obtained via the Model Checking Contest, as well as case studies on distributed systems with other LIP6 teams.

The verification and synthesis of parameterized systems, for instance in the context of asynchronous distributed systems where the communication topology is not fixed, is also a difficult problem as witnessed by the numerous undecidability results. We plan to address theoretical issues and obtain new decidability results in restricted settings, but we also target specific contexts like business processes and networks of mobile robots.

On the more practical side concerning tool development, many cooperations are planned with french and international teams conducting similar activities, in particular in view of extending the expressivity of the GAL language and developing optimization techniques for solving games with imperfect information.

When numerical solutions on stochastic hybrid models do not apply, either because of excessive time or space consumption, or because they require restriction to Markov models, statistical model checking is more adapted. We plan to develop new simulation engines to provide inputs to such tools, for instance to produce performance indices related to the safety of autonomous vehicles (with IRT SystemX) or to handle auto-configuration and auto-adaptation (with UC Labs).

Security issues will be studied in several directions. The first one is the verification of generic properties like opacity or diagnosis for probabilistic models, when an internal agent can either cooperate with an intruder or try to obtain guarantees against attacks. Another direction consists in analyzing information flow by a combination of different approaches, adding static analysis to standard verification techniques, and refining the write/append criterium to obtain better approximations of the flows. Finally, new testing methods will be designed for ensuring data protection in hybrid cloud applications, with an objective of fully automated monitoring.

Future organization of the team. A reflexion to divide the team in two groups, possibly along the two previous scientific objectives, has been initiated and must go on in 2018.

Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
4, place Jussieu - 75005 Paris
http://www.lip6.fr

UPMC SORBONNE UNIVERSITÉS

HCERES
Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur

cnrs