RESEARCH CENTRE
Inria Paris Center

2022
ACTIVITY REPORT

Project-Team
MIMOVE

Middleware on the Move

DOMAIN

**Networks, Systems and Services,
Distributed Computing**

THEME

Distributed Systems and middleware

Innia

Contents

Project-Team MIMOVE

Creation of the Project-Team: 2018 February 01

Keywords

Computer sciences and digital sciences

A1.2.1. – Dynamic reconfiguration

A1.2.3. – Routing

A1.2.4. – QoS, performance evaluation

A1.2.5. – Internet of things

A1.2.6. – Sensor networks

A1.2.7. – Cyber-physical systems

A1.3. – Distributed Systems

A1.4. – Ubiquitous Systems

A1.5. – Complex systems

A1.5.1. – Systems of systems

A1.5.2. – Communicating systems

A2.5. – Software engineering

A2.6.2. – Middleware

A3.1.7. – Open data

A3.1.8. – Big data (production, storage, transfer)

A3.3. – Data and knowledge analysis

A3.5. – Social networks

Other research topics and application domains

B6.3. – Network functions

B6.4. – Internet of things

B6.5. – Information systems

B8.2. – Connected city

B8.5.1. – Participative democracy

1   Team members, visitors, external collaborators

Research Scientists

- Nikolaos Georgantas [Team leader, INRIA, Researcher, HDR]

- Valérie Issarny [INRIA, Senior Researcher, HDR]

Post-Doctoral Fellow

- Maroua Bahri [INRIA]

PhD Students

- Abdoul Shahin Abdoul Soukour [Sorbonne Université]

- William Aboucaya [Sorbonne Université]

- Patient Ntumba Wa Ntumba [INRIA, until May 2022]

Technical Staff

- Zakaria Benomar [INRIA, Engineer, from Jun 2022, Postdoc]

- Patient Ntumba Wa Ntumba [INRIA, Engineer, from Jun 2022, then Postdoc Engineer]

Interns and Apprentices

- Lior Diler [INRIA, Apprentice, until Aug 2022, Engineer]

- Haidong Zhao [INRIA, Intern, from Mar 2022 until Aug 2022, Master's student]

Administrative Assistants

- Nathalie Gaudechoux [INRIA]

- Meriem Guemair [INRIA]

External Collaborators

- Rachit Agarwal [IIT Kanpur, then Merkle Science]

- Rafael Angarita Arocha [ISEP, then Université Paris Nanterre, Associate Professor]

- Georgios Bouloukakis [Télécom SudParis, Associate Professor]

- Vassilis Christophides [ENSEA, Professor]

- Renata Cruz Teixeira [Netflix (secondment from INRIA MIMOVE), Senior Researcher, HDR]

- Bruno Lefèvre [Université Sorbonne Nouvelle]

- Françoise Sailhan [CNAM, then IMT-Atlantique, Professor, HDR]

2   Overall objectives

Given the prevalence of global networking and computing infrastructures (such as the Internet and the Cloud), mobile networking environments, powerful hand-held user devices, and physical-world sensing and actuation devices, the possibilities of new mobile distributed systems have reached unprecedented levels. Such systems are dynamically composed of networked resources in the environment, which may span from the immediate neighborhood of the users – as advocated by pervasive computing – up to the entire globe – as envisioned by the Future Internet and one of its major constituents, the Internet of Things. Hence, we can now talk about truly ubiquitous computing.

The resulting ubiquitous systems have a number of unique – individually or in their combination – features, such as dynamicity due to volatile resources and user mobility, heterogeneity due to constituent resources developed and run independently, and context-dependence due to the highly changing characteristics of the execution environment, whether technical, physical or social. The latter two aspects are particularly manifested through the physical but also social sensing and actuation capabilities of mobile devices and their users. More specifically, leveraging the massive adoption of smart phones and other user-controlled mobile devices, besides physical sensing – where a device's sensor passively reports the sensed phenomena – *social sensing/crowd sensing* comes into play, where the user is aware of and indeed aids in the sensing of the environment.

Mobile systems with the above specifics further push certain problems related to the Internet and user experience to their extreme: (i) Technology is too complex. Most Internet users are not tech-savvy and hence cannot fix performance problems and anomalous network behavior by themselves. The complexity of most Internet applications makes it hard even for networking experts to fully diagnose and fix problems. Users can't even know whether they are getting the Internet performance that they are paying their providers for. (ii) There is too much content. The proliferation of user-generated content (produced anywhere with mobile devices and immediately published in social media) along with the vast amount of information produced by traditional media (e.g., newspapers, television, radio) poses new challenges in achieving an effective, near real-time information awareness and personalization. For instance, users need novel filtering and recommendation tools for helping them to decide which articles to read or which movie to watch.

This challenging context raises key research questions:

- How to deal with heterogeneity and dynamicity, which create runtime uncertainty, when developing and running mobile systems in the open and constantly evolving Internet and IoT environment?

- How to enable automated diagnosis and optimization of networks and systems in the Internet and IoT environment for improving the QoE of their users?

- How to raise human centric crowd-sensing to a reliable means of sensing world phenomena?

- How to deal with combination, analysis and privacy aspects of Web/social media and IoT crowd-sensing data streams?

3   Research program

The research questions identified above call for radically new ways in conceiving, developing and running mobile distributed systems. In response to this challenge, MiMove's research aims at enabling next-generation mobile distributed systems that are the focus of the following research topics.

3.1   Emergent mobile distributed systems

Uncertainty in the execution environment calls for designing mobile distributed systems that are able to run in a beforehand unknown, ever-changing context. Nevertheless, the complexity of such change cannot be tackled at system design-time. Emergent mobile distributed systems are systems which, due to their automated, dynamic, environment-dependent composition and execution, *emerge* in a possibly non-anticipated way and manifest *emergent properties*, i.e., both systems and their properties take their complete form only at runtime and may evolve afterwards. This contrasts with the typical software

engineering process, where a system is finalized during its design phase. MiMove's research focuses on enabling the emergence of mobile distributed systems while assuring that their required properties are met. This objective builds upon pioneering research effort in the area of *emergent middleware* initiated by members of the team and collaborators [3, 5].

3.2 Large-scale mobile sensing and actuation

The extremely large scale and dynamicity expected in future mobile sensing and actuation systems lead to the clear need for algorithms and protocols for addressing the resulting challenges. More specifically, since connected devices will have the capability to sense physical phenomena, perform computations to arrive at decisions based on the sensed data, and drive actuation to change the environment, enabling proper coordination among them will be key to unlocking their true potential. Although similar challenges have been addressed in the domain of networked sensing, including by members of the team [11], the specific challenges arising from the *extremely large scale* of mobile devices – a great number of which will be attached to people, with uncontrolled mobility behavior – are expected to require a significant rethink in this domain. MiMove's research investigates techniques for efficient coordination of future mobile sensing and actuation systems with a special focus on their dependability.

3.3 Mobile social crowd-sensing

While mobile social sensing opens up the ability of sensing phenomena that may be costly or impossible to sense using embedded sensors (e.g., subjective crowdedness causing discomfort or joyfulness, as in a bus or in a concert) and leading to a feeling of being more socially involved for the citizens, there are unique consequent challenges. Specifically, MiMove's research focuses on the problems involved in the combination of the physically sensed data, which are quantitative and objective, with the mostly qualitative and subjective data arising from social sensing. Enabling the latter calls for introducing mechanisms for incentivising user participation and ensuring the privacy of user data, as well as running empirical studies for understanding the complex social behaviors involved. These objectives build upon previous research work by members of the team on mobile social ecosystems and privacy, as well as a number of efforts and collaborations in the domain of smart cities and transport that have resulted in novel mobile applications enabling empirical studies of social sensing systems.

3.4 Active and passive probing methods

We are developing methods that actively introduce probes in the network to discover properties of the connected devices and network segments. We are focusing in particular on methods to discover properties of home networks (connected devices and their types) and to distinguish if performance bottlenecks lie within the home network versus in the different network segments outside (e.g., Internet access provider, interconnects, or content provider). Our goal is to develop adaptive methods that can leverage the collaboration of the set of available devices (including end-user devices and the home router, depending on which devices are running the measurement software).

We are also developing passive methods that simply observe network traffic to infer the performance of networked applications and the location of performance bottlenecks, as well as to extract patterns of web content consumption. We are working on techniques to collect network traffic both at user's end-devices and at home routers. We also have access to network traffic traces collected on a campus network and on a large European broadband access provider.

3.5 Inferring user online experience

We are developing hybrid measurement methods that combine passive network measurement techniques to infer application performance with techniques from HCI to measure user perception as well as methods to directly measure application quality. We later use the resulting datasets to build models of user perception of network performance based only on data that we can obtain automatically from the user device or from user's traffic observed in the network.

3.6 Real time data analytics

The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. The time value of data is crucial for many IoT-based systems requiring *real-time* (or near real-time) *control* and *automation*. Such systems typically collect data continuously produced by "things" (i.e., devices), and analyze them in (sub-) seconds in order to act promptly, e.g., for detecting security breaches of digital systems, for spotting malfunctions of physical assets, for recommending goods and services based on the proximity of potential clients, etc. Hence, they require to both *ingest* and *analyze in real-time* data arriving with different velocity from various IoT data streams.

Existing incremental (online or streaming) techniques for descriptive statistics (e.g., frequency distributions, frequent patterns, etc.) or predictive statistics (e.g., classification, regression) usually assume a good enough quality dataset for mining patterns or training models. However, IoT raw data produced in the wild by sensors embedded in the environment or wearable by users are prone to errors and noise. Effective and efficient algorithms are needed for *detecting* and *repairing data impurities* (for controlling data quality) as well as *understanding data dynamics* (for defining alerts) in real-time, for collections of IoT data streams that might be geographically distributed. Moreover, supervised deep learning and data analytics techniques are challenged by the presence of sparse ground truth data in real IoT applications. Lightweight and adaptive semi-supervised or unsupervised techniques are needed to power real-time anomaly and novelty detection in IoT data streams. The effectiveness of these techniques should be able to reach a useful level through training on a relatively small amount of (preferably unlabeled) data while they can cope distributional characteristics of data evolving over time.

4 Application domains

4.1 Mobile urban systems for smarter cities

With the massive scale adoption of mobile devices and further expected significant growth in relation with the Internet of Things, mobile computing is impacting most – if not all – the ICT application domains. One such domain is the one of "*smart cities*". The smart city vision anticipates that the whole urban space, including buildings, power lines, gas lines, roadways, transport networks, and cell phones, can all be wired together and monitored. Detailed information about the functioning of the city then becomes available to both city dwellers and businesses, thus enabling better understanding and consequently management of the city's infrastructure and resources. This raises the prospect that cities will become more sustainable environments, ultimately enhancing the citizens' well being. There is the further promise of enabling radically new ways of living in, regulating, operating and managing cities, through the increasing active involvement of citizens by ways of crowd-sourcing/sensing and social networking.

Still, the vision of what smart cities should be about has been and keeps evolving at a fast pace in close concert with the latest technology trends. It is notably worth highlighting how mobile and social network use has reignited citizen engagement, thereby opening new perspectives for smart cities beyond data analytics that have been initially one of the core foci for smart cities technologies. Similarly, open data programs foster the engagement of citizens in the city operation and overall contribute to make our cities more sustainable. The unprecedented democratization of urban data fueled by open data channels, social networks and crowd sourcing enables not only the monitoring of the activities of the city but also the assessment of their nuisances based on their impact on the citizens, thereby prompting social and political actions. However, the comprehensive integration of urban data sources for the sake of sustainability remains largely unexplored. This is an application domain that we focus on, further leveraging our research on emergent mobile distributed systems, large-scale mobile sensing & actuation, and mobile social crowd-sensing.

In particular, we concentrate on the following specialized applications:

- **Democratization of urban data for healthy cities.** We integrate the various urban data sources, especially by way of crowd-Xing, to better understand city nuisances. This goes from raw pollution sensing (e.g., sensing noise) to the sensing of its impact on citizens (e.g., how people react to urban noise and how this affects their health).

- **Social applications.** Mobile applications are being considered by sociologists as a major vehicle to actively involve citizens and thereby prompt them to become activists. We study such a vehicle from the ICT perspective and in particular elicit relevant middleware solutions to ease the development of such "*civic apps*".

4.2   Home network diagnosis

With the availability of cheap broadband connectivity, Internet access from the home has become a ubiquity. Modern households host a multitude of networked devices, ranging from personal devices such as laptops and smartphones to printers and media centers. These devices connect among themselves and to the Internet via a local-area network–a home network–that has become an important part of the "Internet experience". In fact, ample anecdotal evidence suggests that the home network can cause a wide array of connectivity impediments, but their nature, prevalence, and significance remain largely unstudied.

Our long-term goal is to assist users with concrete indicators of the quality of their Internet access, causes of potential problems and–ideally–ways to fix them. We intend to develop a set of easy-to-use home network monitoring and diagnosis tools. The development of home network monitoring and diagnosis tools brings a number of challenges. First, home networks are heterogeneous. The set of devices, configurations, and applications in home networks vary significantly from one home to another. We must develop sophisticated techniques that can learn and adapt to any home network as well as to the level of expertise of the user. Second, Internet application and services are also heterogeneous with very diverse network requirements. We must develop methods that can infer application quality solely from the observation of (often encrypted) application network traffic. There are numerous ways in which applications can fail or experience poor performance in home networks. Often there are a number of explanations for a given symptom. We must devise techniques that can identify the most likely cause(s) for a given problem from a set of possible causes. Finally, even if we can identify the cause of the problem, we must then be able to identify a solution. It is important that the output of the diagnosis tools we build is "actionable". Users should understand the output and know what to do.

In our patternship with Princeton University (associate team HOMENET) we have deployed monitoring infrastructure within users' homes. We are developing a mostly passive measurement system to monitor the performance of user applications, which we call Network Microscope. We are developing Network Microscope to run in a box acting as home gateway. We have deployed these boxes in 50 homes in the US and 10 in France. The US deployment was ran and financed by the Wall Street Journal. They were interested in understanding the relationship between Internet access speed and video quality. We have been discussing with Internet regulators (in particular, FCC, ACERP, and BEREC) as well as residential access ISP in how Network Microscope can help overcome the shortcomings of existing Internet quality monitoring systems.

4.3   Mobile Internet quality of experience

Mobile Internet usage has boomed with the advent of ever smarter handheld devices and the spread of fast wireless access. People rely on mobile Internet for everyday tasks such as banking, shopping, or entertainment. The importance of mobile Internet in our lives raises people's expectations. Ensuring good Internet user experience (or Quality of Experience—QoE) is challenging, due to the heavily distributed nature of Internet services. For mobile applications, this goal is even more challenging as access connectivity is less predictable due to user mobility, and the form factor of mobile devices limits the presentation of content. For these reasons, the ability to monitor QoE metrics of mobile applications is essential to determine when the perceived application quality degrades and what causes this degradation in the chain of delivery. Our goal is to improve QoE of mobile applications.

To achieve this goal, we are working on three main scientific objectives. First, we are working on novel methods to monitor mobile QoE. Within the IPL BetterNet we are developing the HostView for Android tool that runs directly on mobile devices to monitor network and system performance together with the user perception of performance. Second, we plan to develop models to predict QoE of mobile applications. We will leverage the datasets collected with HostView for Android to build data-driven models. Finally, our goal is to develop methods to optimize QoE for mobile users. We are currently developing optimization

methods for interactive video applications. We envision users walking or driving by road-side WiFi access points (APs) with full 3G/LTE coverage and patchy WiFi coverage (i.e., community Wifi or Wifi APs on Lampposts) or devices with multiple 3G/LTE links. To achieve this goal, we plan to leverage multi-path and cross-layer optimizations.

4.4 Internet Scanning

Internet-wide scanning has enabled researchers to answer a wealth of new security and measurement questions ranging from "How are authoritarian regimes spying on journalists?" to "Are security notifications effective at prompting operators to patch?" Most of these studies have used tools like ZMap, which operates naively, scanning every IPv4 address once. This simplicity enables researchers to easily answer a question once, but the methodology scales poorly when continually scanning to detect changes, as networks change at dramatically different rates. Service configurations change more frequently on cloud providers like Amazon and Azure than on residential networks. Internet providers in developing regions often have extremely short DHCP windows. Some networks are unstable with host presence varying wildly between different hours and others have distinct periodic patterns, e.g., hosts are only available during regional business hours. A handful of large autonomous systems have not had hosts present in decades. Our work in collaboration with Stanford University is developing more intelligent Internet-wide scanning methods to then implement a system that can scan continuously. Such a system will allow for up-to-date analysis of Internet trends and threats with real-time alerts of important events.

5 Highlights of the year

Valérie Issarny, founder of the ARLES project-team (predecessor of MIMOVE) and cornerstone member of MIMOVE, passed on November 12, 2022. The team continues Valérie's combat, which she carried out passionately until the very end, for free academic research as the principal mission of Inria.

Other highlights:

- Valérie Issarny received the 2022 IEEE Computer Society Technical Committee on Software Engineering "Distinguished Service" Award in recognition for significant contributions in service, mentorship, and influence to the Software Engineering community.

- Valérie Issarny is General Co-Chair of the 8th International Conference on Smart Computing (SMARTCOMP), Helsinki, Finland, June 2022.

- Valérie Issarny is TPC Co-Chair of the 18th Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), Melbourne, Australia, May 2023.

6 New software and platforms

6.1 New software

6.1.1 DeXMS

Name: Data eXchange Mediator Synthesizer

Keywords: Internet of things, Middleware protocol interoperability, Edge Computing

Functional Description: To deal with the high technology diversity of the IoT solutions landscape, we have introduced a systematic solution to the IoT interoperability problem at the middleware layer. We identify common interaction abstractions across the multitude of existing heterogeneous IoT protocols and model them into the DeX (Data eXchange) API & connector model. We further elicit the DeXIDL (Interface Description) language to describe the application interfaces of Things in a common abstract way. Based on DeX and DeXIDL, we introduce an architecture for mediators that can bridge heterogeneous Things and their protocols. The outcome of our overall effort is the

DeXMS (Mediator Synthesizer) development & runtime framework, which supports the automated synthesis, deployment and execution of mediators at the edge.

URL: https://gitlab.inria.fr/DeXMS

Contact: Nikolaos Georgantas

Participants: Georgios Bouloukakis, Nikolaos Georgantas, Patient Ntumba Wa Ntumba

6.1.2 SenseTogether

Keywords: Mobile Crowdsensing, Sensor Calibration, Context Inference, Edge Computing

Functional Description: Our work aims to raise opportunistic mobile crowdsensing to a reliable means of observing phenomena, focusing on urban environmental monitoring. More specifically, the mobile crowdsensors contribute measurements related to the physical environment (e.g., ambient temperature, air pressure, ambient humidity, ambient light, sound level, magnetic field) using the embedded/connected sensors on smart devices. To this end, we have developed a set of protocols that together support "context-aware collaborative mobile crowdsensing at the edge", by combining the following complementary features:

(i) CalibrateNoiseTogether: Multi-hop, multiparty calibration to ensure the accuracy of sensors embedded in or connected to smartphones. Sensors that are within a relevant sensing and communication range coordinate so that the observations of previously calibrated sensors serve calibrating new sensors.

(ii) ContextSense: Inference of the crowdsensors' physical context so as to characterize the gathered data. Indeed, the relevance of the provided measurements depends on the adequacy of the sensing context with respect to the analyzed phenomena. We introduce an online learning approach to support the local inference of the sensing context that can evolve according to the environment in which it takes place.

(iii) BeTogether: Context-aware grouping of crowdsensors to share the workload and filter out low quality data. We leverage D2D communication and introduce a context-aware and cloudless collaboration strategy in which crowdsensor groups are maintained in an autonomous and distributed way to monitor a physical phenomenon of interest.

(iv) IAM (Interpolation and Aggregation on the Move): Data processing at the edge to enhance the knowledge transferred to the cloud and reduce the data uploading and resource consumption in the cloud. The data interpolation and aggregation is based on opportunistic meetings of the crowdsensors, and the relay decision is made based on the quality of the inferred data.

URL: https://github.com/sensetogether

Contact: Valerie Issarny

Participants: Yifan Du, Françoise Sailhan, Valerie Issarny

6.1.3 SocialBus

Name: Universal Social Network Bus

Keywords: Middleware, Interoperability, Social networks, Software Oriented Service (SOA)

Functional Description: Online social network services (OSNSs) have become an integral part of our daily lives. At the same time, the aggressive market competition has led to the emergence of multiple competing siloed OSNSs that cannot interoperate. As a consequence, people face the burden of creating and managing multiple OSNS accounts and learning how to use them, to stay connected. The goal of the Universal Social Network Bus (USNB) is to relieve users from such a burden, letting them use their favorite applications to communicate.

URL: http://cicamo.re/#socialbus

Contact: Valerie Issarny

Participants: Rafael Angarita Arocha, Lior Diler, William Aboucaya, Valerie Issarny, Nikolaos Georgantas

6.1.4   Network Microscope

Keywords: Quality of Experience, Network monitoring, Video analysis

Functional Description: A system that accurately infers video streaming quality metrics in real time, such as startup delay or video resolution, by using just a handful of features extracted from passive traffic measurement. Network Microscope passively collects a corpus of network features about the traffic flows of interest in the network and directs those to a real-time analytics framework that can perform more complex inference tasks. Network Microscope enables network operators to determine degradations in application quality as they happen, even when the traffic is encrypted.

URL: https://netmicroscope.com/

Contact: Renata Cruz Teixeira

Participants: Francesco Bronzino, Renata Cruz Teixeira

7   New results

7.1   Scheduling of Continuous Operators for IoT Edge Analytics with Time Constraints

Participants: Patient Ntumba (MiMove), Nikolaos Georgantas (MiMove), Vassilis Christophides (ENSEA)

Data stream processing and analytics (DSPA) engines are used to extract in (near) real-time valuable information from multiple IoT data streams. Deploying DSPA applications at the IoT network edge through Edge/Fog architectures is currently one of the core challenges for reducing both network delays and network bandwidth usage to reach the Cloud. In this paper, we address the problem of scheduling continuous DSPA operators to Fog-Cloud nodes featuring both computational and network resources. We are paying particular attention to the dynamic workload of these nodes due to variability of IoT data stream rates and the sharing of nodes' resources by multiple DSPA applications. In this respect, we propose TSOO, a resource-aware and time-efficient heuristic algorithm that takes into account the limited Fog computational resources, the real-time response constraints of DSPA applications, as well as, congestion and delay issues on Fog-to-Cloud network resources. Via extensive simulation experiments, we show that TSOO approximates an optimal operators' placement with a low execution cost.

7.2   Supporting Multi-Cloud in Serverless Computing

Participants: Haidong Zhao (MiMove & Sorbonne Univesité & Technische Universitat Berlin), Zakaria Benomar (MiMove), Tobias Pfandzelter (Technische Universitat Berlin), Nikolaos Georgantas (MiMove)

Serverless computing is a widely adopted cloud execution model composed of Function-as-a-Service (FaaS) and Backend-as-a-Service (BaaS) offerings. The increased level of abstraction makes vendor lock-in inherent to serverless computing, raising more concerns than previous cloud paradigms. Multicloud serverless is a promising emerging approach against vendor lock-in, yet multiple challenges must be overcome to tap its potential. First, we need to be aware of both the performance and cost of each FaaS provider. Second, a multi-cloud architecture needs to be proposed before deploying a multi-cloud workflow. Domain-specific serverless offerings must then be integrated into the multi-cloud architecture to improve performance and/or save costs. Finally, we require workload portability support for serverless multi-cloud. In this work, we present a multi-cloud library for cross-serverless offerings. We develop an analysis system to support comparison among public FaaS providers in terms of performance and cost. Moreover, we present how to alleviate data gravity with domain-specific serverless offerings. Finally, we deploy workloads on these architectures to evaluate several public FaaS offerings.

7.3   Consent-driven Data Reuse in Multi-tasking Crowdsensing Systems: A Privacy-by-Design Solution

Participants: Mariem Brahem (Inria PETRUS & UVSQ DAVID), Guillaume Scerri (Inria PETRUS & UVSQ DAVID), Nicolas Anciaux (Inria PETRUS & UVSQ DAVID), Valérie Issarny (MiMove)

Mobile crowdsensing allows gathering massive data across time and space to feed our environmental knowledge, and to link such knowledge to user behavior. However, a major challenge facing mobile crowdsensing is to guarantee privacy preservation to the contributing users. Privacy preservation in crowdsensing systems has led to two main approaches, sometimes combined, which are, respectively, to trade privacy for rewards, and to take advantage of privacy-enhancing technologies "anonymizing" the collected data. Although relevant, we claim that these approaches do not sufficiently take into account the users' own tolerance to the use of the data provided, so that the crowdsensing system guarantees users the expected level of confidentiality as well as fosters the use of crowdsensing data for different tasks. To this end, we leverage the l-Completeness property, which ensures that the data provided can be used for all the tasks to which their owners consent as long as they are analyzed with l-1 other sources, and that no privacy violations can occur due to the related contribution of users with less stringent privacy requirements. The challenge, therefore, is to ensure l-Completeness when analyzing the data while allowing the data to be used for as many tasks as possible, and promoting the accuracy of the resulting knowledge. This is achieved through a clustering algorithm sensitive to the data distribution, which optimizes data reuse and utility. Nevertheless, it is critical to allow the deployment of such a solution even in the presence of a malicious adversary able to act on the server side, for which we introduce a privacy-by-design architecture leveraging Trusted Execution Environments. The implementation of a prototype using SGX enclaves further allows running experiments that show that our system incurs a reasonable performance overhead, while providing strong security properties against a malicious adversary.

7.4   Detecting Obstacles to Collaboration in an Online Participatory Democracy Platform: A Use-case Driven Analysis

Participants: William Aboucaya (MiMove), Rafael Angarita (LISITE - ISEP), Valérie Issarny (MiMove)

Massive online participatory platforms are an essential tool for involving citizens in public decision-making on a large scale, both in terms of the number of participating citizens and their geographical distribution. However, engaging a sufficiently large number of citizens, as well as collecting adequate contributions, require special attention in the functionalities implemented by the platform. This paper empirically analyzes the existing flaws in participatory platforms and their impact on citizen participation. We focus specifically on the citizen consultation "République Numérique" (Digital Republic) to identify issues arising from the interactions between users on the supporting platform. We chose this consultation because of the high number of contributors and contributions, and the various means of interaction it proposes. Through an analysis of the available data, we highlight that contributions tend to be concentrated around a small set of proposals and contributors. This leads us to formulate a number of recommendations for the design of participatory platforms regarding the management of contributions, from their organization to their presentation to users.

7.5   Effective Weighted k-Nearest Neighbors for Dynamic Data Streams

Participants: Maroua Bahri (MiMove)

Many real-world applications involve classification from evolving data streams. However, learning in such environment requires algorithms able to learn and predict from potentially unbounded data that are constantly changing. For this to happen, stream algorithms should restrict the storage to a part of – and/or synopsis information from – the stream using efficient and accurate manners and strategies, such as window models and summarization techniques (e.g., sampling, sketching, dimensionality reduction). In this work, we focus on the k-Nearest Neighbors (kNN) where most of the existing approaches for data streams consider that instances have the same weight from the start to the finish of the processing task.

In a streaming data scenario, it is often the case that the most recent elements from the data stream are the more relevant ones. Taking into account that the most recent instances are more relevant, we propose a novel kNN approach that stores instances in a sliding window and weighs them according to their arrival time (i.e position on the window) using an adjusted weight function. The empirical results on comprehensive real and synthetic datasets indicate the effectiveness and efficiency of our proposed approach in comparison with state-of-the-art algorithms.

7.6   AutoAD: an Automated Framework for Unsupervised Anomaly Detection

Participants: Andrian Putina (LTCI, Télécom Paris), Maroua Bahri (MiMove), Flavia Salutari (LTCI, Télécom Paris), Mauro Sozio (LTCI, Télécom Paris)

Over the last decade, we witnessed the proliferation of several machine learning algorithms capable of solving different tasks for the most diverse applications. Often, for an algorithm to be effective, significant human effort is required, in particular for hyper-parameter tuning and data cleaning. Recently, there have been increasing efforts to alleviate such a burden and make machine learning algorithms easier to use for researchers with varying levels of expertise. Nevertheless, the question of whether an efficient and fully generalizable automated Machine Learning (autoML) framework is possible remains unanswered. In this work, we present autoAD, the first autoML framework for unsupervised anomaly detection. By leveraging a pool of different anomaly detection algorithms, each one coming with its own hyper-parameter search space, our framework automatically selects the best performing approach, while determining an optimal configuration for its hyperparameters on a given dataset. Our extensive experimental evaluation, conducted on a rich collection of datasets, shows the substantial gains that can be achieved with autoAD compared to state-of-the-art methods for unsupervised anomaly detection.

7.7   Evolution-based Online Automated Machine Learning

Participants: Cedric Kulbach (FZI), Jacob Montiel (University of Waikato), Maroua Bahri (MiMove), Marco Heyden (FZI), and Albert Bifet (University of Waikato)

Automated Machine Learning (AutoML) deals with finding well-performing machine learning models and their corresponding configurations without the need of machine learning experts. However, if one assumes an online learning scenario, where an AutoML instance executes on evolving data streams, the question for the best model and its configuration with respect to occurring changes in the data distribution remains open. Algorithms developed for online learning settings rely on few and homogeneous models and do not consider data mining pipelines or the adaption of their configuration. We, therefore, introduce EvoAutoML, an evolution-based online learning framework consisting of heterogeneous and connectable models that supports large and diverse configuration spaces and adapts to the online learning scenario. We present experiments with an implementation of EvoAutoML on a diverse set of synthetic and real datasets, and show that our proposed approach outperforms state-of-the-art online algorithms as well as strong ensemble baselines in a traditional test-then-train evaluation.

7.8   AutoML: state of the art with a focus on anomaly detection, challenges, and research directions

Participants: Maroua Bahri (MiMove), Flavia Salutari (LTCI, Télécom Paris), Andrian Putina (LTCI, Télécom Paris), Mauro Sozio (LTCI, Télécom Paris)

The last decade has witnessed the explosion of machine learning research studies with the inception of several algorithms proposed and successfully adopted in different application domains. However, the performance of multiple machine learning algorithms is very sensitive to multiple ingredients (e.g., hyper-parameters tuning and data cleaning) where a significant human effort is required to achieve good results. Thus, building well-performing machine learning algorithms requires domain knowledge and highly specialized data scientists. Automated machine learning (autoML) aims to make easier and more accessible the use of machine learning algorithms for researchers with varying levels of expertise.

Besides, research effort to date has mainly been devoted to autoML for supervised learning, and only a few research proposals have been provided for the unsupervised learning. In this work, we present an overview of the autoML field with a particular emphasis on the automated methods and strategies that have been proposed for unsupervised anomaly detection.

7.9 Predicting IPv4 services across all ports

Participants: Liz Izhikevich (Stanford University), Renata Teixeira (MiMove), Zakir Durumeric (Stanford University)

Internet-wide scanning is commonly used to understand the topology and security of the Internet. However, IPv4 Internet scans have been limited to scanning only a subset of services—exhaustively scanning all IPv4 services is too costly and no existing bandwidth-saving frameworks are designed to scan IPv4 addresses across all ports. In this work we introduce GPS, a system that efficiently discovers Internet services across all ports. GPS runs a predictive framework that learns from extremely small sample sizes and is highly parallelizable, allowing it to quickly find patterns between services across all 65K ports and a myriad of features. GPS computes service predictions in 13 minutes (four orders of magnitude faster than prior work) and finds 92.5% of services across all ports with 131× less bandwidth, and 204× more precision, compared to exhaustive scanning. GPS is the first work to show that, given at least two responsive IP addresses on a port to train from, predicting the majority of services across all ports is possible and practical.

8 Bilateral contracts and grants with industry

8.1 Bilateral grants with industry

- "Monitoring and diagnosis of Internet QoE", Google Faculty Award to Renata Teixeira and D. Choffnes (Northeastern University), 2017-2022.

- "Application Performance Bottleneck Detection", Comcast Gift to Renata Teixeira, 2018-2022.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Associate Teams in the framework of an Inria International Lab or in the framework of an Inria International Program

MINES

Title: Adaptive Communication Middleware for Resilient Sensing & Actuation IN Emergency Response Scenarios

Partner Institutions:

- Distributed Systems Middleware (DSM) group, University of California, Irvine (Nalini Venkata-subramanian)
- Inria MiMove (Valérie Issarny)

Duration: 2018 - 2022

URL: MINES

Additional info: Emerging smart-city and smart-community efforts will require a massive deployment of connected entities (Things) to create focused smartspaces. Related applications will enhance citizen quality of life and public safety (e.g., providing safe evacuation routes in fires). However, supporting IoT deployments are heterogeneous and can be volatile and failure-prone as they are

often built upon low-powered, mobile and inexpensive devices - the presence of faulty compo-
nents and intermittent network connectivity, especially in emergency scenarios, tend to deliver
inaccurate/delayed information. The MINES associate team addresses the resulting challenge of
enabling interoperability and resilience in large-scale IoT systems through the design and develop-
ment of a dedicated middleware. More specifically, focusing on emergency situations, the MINES
middleware will: (i) enable the dynamic composition of IoT systems from any and all available
heterogeneous devices; (ii) support the timely and reliable exchange of critical data within and
across IoT in the enabled large-scale and dynamic system over heterogeneous networks. Finally,
the team will evaluate the proposed solution in the context of emergency response scenario use
cases.

9.2   European initiatives

9.2.1   Horizon Europe

SEDIMARK

> **Participants:**   Nikolaos Georgantas, Valérie Issarny.

SEDIMARK project on cordis.europa.eu

Title: SEcure Decentralised Intelligent Data MARKetplace

Duration: From October 1, 2022 to September 30, 2025

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- WINGS ICT SOLUTIONS INFORMATION & COMMUNICATION TECHNOLOGIES IKE, Greece
- UNIVERSITY COLLEGE DUBLIN, NATIONAL UNIVERSITY OF IRELAND, DUBLIN (NUID UCD), Ireland
- FORUM VIRIUM HELSINKI OY (RADIO- JATELEVISIOTEKNIIKAN TUTKIMUS RTT), Finland
- SIEMENS SRL, Romania
- ATOS SPAIN SA, Spain
- AYUNTAMIENTO DE SANTANDER (AYTO SANTANDER), Spain
- MYTILINAIOS ANONIMI ETAIREIA (MYTlLINEOS), Greece
- UNIVERSIDAD DE CANTABRIA (UC), Spain
- FONDAZIONE LINKS - LEADING INNOVATION & KNOWLEDGE FOR SOCIETY (FONDAZIONE LINKS), Italy
- ATOS IT SOLUTIONS AND SERVICES IBERIA SL (ATOS IT), Spain
- UNIVERSITY OF SURREY (SURREY), United Kingdom
- EASY GLOBAL MARKET SAS (EGLOBALMARK), France

Inria contact: Nikolaos Georgantas

Coordinator: ATOS SPAIN SA, Spain

Summary: The EU data economy has grown tremendously, with forecasts predicting to reach 800Billion
Euros in 2025. Data are becoming the new currency, being exchanged as products or services in
marketplaces. Data markets are predicted to reach a size of 100Billion Euros in 2025. Existing
data marketplaces are centralised, store the data on the cloud, provide limited to no guarantees
about data quality and they are governed by single entities that make the rules. SEDIMARK merges

the expertise of a large team of experts to build a secure, trusted and intelligent decentralised data and services marketplace, based on Distributed Ledger Technology and Artificial Intelligence. SEDIMARK enables distributed heterogeneous data within the EU to be easily and seamlessly linked, shared and exploited for diverse business and research scenarios. SEDIMARK builds upon the concept of FAIR data, ensuring that data are of the highest quality, unbiased, enriched and annotated, so that they can be discovered, accessed, and easily reused. SEDIMARK includes a distributed registry of resources (data/services) stored on edge systems, close to where they are generated and where the data are cleaned, labelled, validated and anonymised. Security is applied with strong access control, privacy techniques for data minimisation and purpose limitation, exploiting blockchain for enforcing trust, decentralised identities, and data verification. Energy efficient AI techniques will be used for automated data quality management, labelling and classification of data as well as for providing (distributed) analytics and advanced services on top of the data. Semantic interoperability based on common ontologies and data models will allow the easy and efficient discovery, sharing and federation of heterogeneous data from multiple sources. The system is built on top of existing platforms of the consortium, starting from TRL5 and will be tested and demonstrated in four real world scenarios, reaching TRL-8.

9.3   National initiatives

BPI – France Relance – 5G Events Labs

> **Participants:**   Nikolaos Georgantas, Patient Ntumba, Zakaria Benomar.

Partner Institutions:

- Orange
- Ericsson
- Inria
- CEA - Centre de Saclay

Duration: 2021 - 2023

Additional info: The 5G Events Labs project aims to boost the economic activity of the events, culture and sports sectors, around ten major sites in France where Orange and its partners will offer 5G coverage, technological platforms and adapted support enabling companies to leverage these technologies and incubate innovations in the areas of services for attendees and organizers. MIMOVE brings expertise in middleware solutions for the IoT that support intelligent spaces and applications across the mobile-edge-cloud continuum.

10   Dissemination

10.1   Promoting scientific activities

10.1.1   Scientific events: organisation

General chair, scientific chair

- V. Issarny, General Co-Chair of the 8th IEEE International Conference on Smart Computing (SMART-COMP 2022), Helsinki, Finland, June 2022.

10.1.2   Scientific events: selection

Chair of conference program committees

- V. Issarny, TPC Co-Chair of the 18th Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), Melbourne, Australia, May 2023.

Member of the conference program committees

- V. Issarny, member of the TPC of the following international conferences: IEEE ICDCS'22, ACM/IFIP Middleware'23.

- N. Georgantas, member of the TPC of the following international conferences: ACM SAC'22, IEEE SOSE'22, The ACM Web Conference'22, IEEE WETICE'22, IEEE SMARTCOMP'22, SEAMS'22, IEEE MetaCom'23.

- N. Georgantas, member of the TPC of the following international workshop: SERENE'22.

10.1.3 Journal

Member of the editorial boards

- V. Issarny, appointed Editor-in-Chief of ACM Transactions on Autonomous and Adaptive Systems (TAAS).

- V. Issarny, member of the following editorial boards: ACM Transactions on the Internet of Things (TIOT); IEEE Transactions on Services Computing (TSC); IEEE Transactions on Software Engineering (TSE).

10.1.4 Leadership within the scientific community

- V. Issarny, Elected Chair of ACM Europe Council, 2021-2023.

10.1.5 Scientific expertise

- V. Issarny, member of: Inria Evaluation Committee (Elected, 2019-2022), FWO expert panel for PhD fellowships on strategic basic research (appointed since 2018).

10.2 Teaching - Supervision - Juries

10.2.1 Supervision

- PhD: Patient Ntumba, "Scheduling Streaming Operators for IoT Edge Analytics", Sorbonne University, September 9, 2022, N. Georgantas and V. Christophides (ENSEA).

- PhDs in progress:

  - William Aboucaya (from October 2019): "Version control for urban participatory systems", Sorbonne University, V. Issarny and R. Angarita (ISEP).

  - Abdoul Shahin Abdoul Soukour (from October 2020): "Goal-driven automated composition of Function-as-a-Service workflows", Sorbonne University, N. Georgantas.

11 Scientific production

11.1 Major publications

[1] R. Almeida, Í. Cunha, R. Teixeira, D. Veitch and C. Diot. 'Classification of Load Balancing in the Internet'. In: *IEEE INFOCOM 2020 - International Conference on Computer Communications*. Beijing / Virtual, China, Apr. 2020. URL: https://hal.inria.fr/hal-02396406.

[2] R. Angarita, B. Lefèvre, S. Ahvar, E. Ahvar, N. Georgantas and V. Issarny. 'Universal Social Network Bus: Towards the Federation of Heterogeneous Online Social Network Services'. In: *ACM Transactions on Internet Technology* (2019). DOI: 10.1145/3323333. URL: https://hal.inria.fr/hal-02072544.

[3]   A. Bennaceur and V. Issarny. 'Automated Synthesis of Mediators to Support Component Interoperability'. In: *IEEE Transactions on Software Engineering* (2015), p. 22. URL: https://hal.inria.fr/hal-01076176.

[4]   B. Billet and V. Issarny. 'Spinel: An Opportunistic Proxy for Connecting Sensors to the Internet of Things'. In: *ACM Transactions on Internet Technology* 17.2 (Mar. 2017), pp. 1–21. DOI: 10.1145/3041025. URL: https://hal.inria.fr/hal-01505879.

[5]   G. Blair, A. Bennaceur, N. Georgantas, P. Grace, V. Issarny, V. Nundloll and M. Paolucci. 'The Role of Ontologies in Emergent Middleware: Supporting Interoperability in Complex Distributed Systems'. In: *Big Ideas track of ACM/IFIP/USENIX 12th International Middleware Conference*. Lisbon, Portugal, 2011. URL: http://hal.inria.fr/inria-00629059/en.

[6]   G. Bouloukakis, N. Georgantas, P. Ntumba and V. Issarny. 'Automated synthesis of mediators for middleware-layer protocol interoperability in the IoT'. In: *Future Generation Computer Systems* 101 (Dec. 2019), pp. 1271–1294. DOI: 10.1016/j.future.2019.05.064. URL: https://hal.inria.fr/hal-02304074.

[7]   F. Bronzino, P. Schmitt, S. Ayoubi, G. Martins, R. Teixeira and N. Feamster. 'Inferring Streaming Video Quality from Encrypted Traffic: Practical Models and Deployment Experience'. In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems*. ACM Sigmetrics 2020 3.3 (Dec. 2019). DOI: 10.1145/3366704. URL: https://hal.inria.fr/hal-03144919.

[8]   M. Caporuscio, P.-G. Raverdy and V. Issarny. 'ubiSOAP: A Service Oriented Middleware for Ubiquitous Networking'. In: *IEEE Transactions on Services Computing* 99 (2012). DOI: \url{http://doi.ieeecomputersociety.org/10.1109/TSC.2010.60}. URL: http://hal.inria.fr/inria-00519577.

[9]   Y. Du, F. Sailhan and V. Issarny. 'Let Opportunistic Crowdsensors Work Together for Resource-efficient, Quality-aware Observations'. In: *PerCom 2020: IEEE International Conference on Pervasive Computing and Communications*. Austin / Virtual, United States, Mar. 2020. DOI: 10.1109/PerCom45495.2020.9127391. URL: https://hal.archives-ouvertes.fr/hal-02463610.

[10]   G. Grassi, R. Teixeira, C. Barakat and M. Crovella. 'Leveraging Website Popularity Differences to Identify Performance Anomalies'. In: *INFOCOM 2021 - IEEE International Conference on Computer Communications*. Vancouver / Virtual, Canada, May 2021. URL: https://hal.inria.fr/hal-03109717.

[11]   S. Hachem, A. Pathak and V. Issarny. 'Service-Oriented Middleware for Large-Scale Mobile Participatory Sensing'. In: *Pervasive and Mobile Computing* (2014). URL: http://hal.inria.fr/hal-00872407.

11.2   Publications of the year

International journals

[12]   M. Bahri, F. Salutari, A. Putina and M. Sozio. 'AutoML: state of the art with a focus on anomaly detection, challenges, and research directions'. In: *International Journal of Data Science and Analytics* (4th Feb. 2022). DOI: 10.1007/s41060-022-00309-0. URL: https://hal.inria.fr/hal-03590242.

[13]   M. Brahem, G. Scerri, N. Anciaux and V. Issarny. 'Consent-driven Data Reuse in Multi-tasking Crowdsensing Systems: A Privacy-by-Design Solution'. In: *Pervasive and Mobile Computing* 83 (2022). DOI: 10.1016/j.pmcj.2022.101614. URL: https://hal.archives-ouvertes.fr/hal-03775759.

International peer-reviewed conferences

[14] W. Aboucaya, R. Angarita and V. Issarny. 'Detecting Obstacles to Collaboration in an Online Participatory Democracy Platform: A Use-case Driven Analysis'. In: FairWare '22 - International Workshop on Equitable Data and Technology in conjunction with the 44th International Conference on Software Engineering (ICSE 2022). Pittsburgh, PA, United States: Association for Computing Machinery, 9th May 2022, pp. 25–33. DOI: 10.1145/3524491.3527307. URL: https://hal.inria.fr/hal-03654027.

[15] M. Bahri. 'Effective Weighted k-Nearest Neighbors for Dynamic Data Streams'. In: 7th Workshop on Real-time Stream Analytics, Stream Mining, CER/CEP & Stream Data Management in Big Data in conjunction with the IEEE International Conference on Big Data 2022. Osaka, Japan, 22nd Dec. 2022. URL: https://hal.inria.fr/hal-03938177.

[16] L. Izhikevich, R. Teixeira and Z. Durumeric. 'Predicting IPv4 services across all ports'. In: SIGCOMM '22: ACM SIGCOMM 2022 Conference. Amsterdam, Netherlands: ACM, 22nd Aug. 2022, pp. 503–515. DOI: 10.1145/3544216.3544249. URL: https://hal.inria.fr/hal-03790076.

[17] C. Kulbach, J. Montiel, M. Bahri, M. Heyden and A. Bifet. 'Evolution-Based Online Automated Machine Learning'. In: PAKDD 2022 - Pacific-Asia Conference on Knowledge Discovery and Data Mining. Vol. 13280. Lecture Notes in Computer Science. Chengdu, China: Springer International Publishing, 10th May 2022, pp. 472–484. DOI: 10.1007/978-3-031-05933-9_37. URL: https://hal.inria.fr/hal-03667231.

[18] P. Ntumba, V. Christophides and N. Georgantas. 'Scheduling of Continuous Operators for IoT edge Analytics with Time Constraints'. In: SMARTCOMP 2022 - International Conference on Smart Computing. Espoo, Finland, 20th June 2022. URL: https://hal.inria.fr/hal-03701939.

[19] A. Putina, M. Bahri, F. Salutari and M. Sozio. 'AutoAD: an Automated Framework for Unsupervised Anomaly Detection'. In: DSAA 2022 - IEEE International Conference on Data Science and Advanced Analytics. Paris / Virtual Event, France: IEEE, 13th Oct. 2022. URL: https://hal.inria.fr/hal-03811809.

[20] H. Zhao, Z. Benomar, T. Pfandzelter and N. Georgantas. 'Supporting Multi-Cloud in Serverless Computing'. In: CloudAM: 11th International Workshop on Cloud and Edge Computing, and Applications Management in conjunction with the 15th IEEE/ACM Utility and Cloud Computing Conference (UCC). Vancouver, United States, 6th Dec. 2022. URL: https://hal.inria.fr/hal-03945892.

Doctoral dissertations and habilitation theses

[21] P. Ntumba wa Ntumba. 'Scheduling Streaming Operators for IoT Edge Analytics'. Sorbonne Université, 9th Sept. 2022. URL: https://theses.hal.science/tel-03846698.

Reports & preprints

[22] K. Kaur, F. Guillemin and F. Sailhan. *Container placement and migration strategies for Cloud, Fog and Edge data centers: A survey.* 12th Apr. 2022. URL: https://hal.archives-ouvertes.fr/hal-03638246.

Other scientific publications

[23] H. Zhao. 'Managing Vendor Lock-in in Serverless Edge-to-Cloud Computing from the Client Side'. Technical University of Berlin, 2nd Dec. 2022. URL: https://hal.inria.fr/hal-03946722.