RESEARCH CENTRE

Paris

2021

ACTIVITY REPORT

Project-Team

MIMOVE

Middleware on the Move

DOMAIN

Networks, Systems and Services, Distributed Computing

THEME

Distributed Systems and middleware

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, from Oct 2021]

PhD Students

- Abdoul Shahin Abdoul Soukour [Sorbonne Université]
- William Aboucaya [Sorbonne Université]
- Patient Ntumba [Inria]

Technical Staff

- Pierre-Guillaume Raverdy [Inria, Engineer, until Mar 2021]

Interns and Apprentices

- Lior Diler [Inria, Apprentice]

Administrative Assistants

- Nathalie Gaudechoux [Inria]
- Meriem Guemair [Inria]

External Collaborators

- Rachit Agarwal [IIT Kanpur]
- Rafael Angarita Arocha [ISEP]
- Sara Ayoubi [Nokia Bell Labs]
- Georgios Bouloukakis [Télécom SudParis]
- Francesco Bronzino [Nokia Bell Labs, then Université Savoie Mont Blanc]
- Vassilis Christophides [ENSEA]
- Renata Cruz Teixeira [Netflix (secondment from Inria MiMove)]
- Yifan Du [Huawei]
- Simin Ghasemi Falavarjani [Université d'Ispahan]
- Giulio Grassi [CISCO Systems]
- Bruno Lefèvre [Université Sorbonne Nouvelle]
- Françoise Sailhan [CNAM]

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 naiıvely, 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

- V. Issarny is Elected Chair of ACM Europe Council, 2021-2023.
- V. Issarny is general co-chair of 2021 IEEE International Conference on Services Computing (SCC) as part of the 2021 IEEE World Congress on Services (SERVICES), Virtual, September 2021.
- V. Issarny is general co-chair of 6th ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI), CPS-IOT week, Virtual, May 2021.
- V. Issarny is TPC co-chair of 7th International Conference on Smart Computing (SMARTCOMP), Virtual, June 2021.
- R. Teixeira is TPC co-chair of 18th USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2021.

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

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 Continuous Operators for IoT Edge Analytics

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

In this work, we are interested in exploring the Edge-Fog-Cloud architecture as an alternative approach to the Cloud-based IoT data analytics. Given the limitations of Fog in terms of limited computational resources that can also be shared among multiple analytics with continuous operators over data streams, we introduce a holistic cost model that accounts both the network and computational resources available in the Edge-Fog-Cloud architecture. Then, we propose scheduling algorithms RCS and SOO-CPLEX for placing continuous operators for data stream analytics at the network edge. The former dynamically places continuous operators between the Cloud and the Fog according to the evolution of data streams rates and uses as less as possible Fog computational resources to satisfy the constraints regarding the usage of both computational and network resources. The latter statically places continuous operators between the Cloud and the Fog to minimize the overall computational and network resource usage cost. Based on thorough experiments, we evaluate the effectiveness of SOO-CPLEX and RCS using simulation.

7.2 Efficient Scheduling of Streaming Operators for IoT Edge Analytics

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

Data stream processing and analytics (DSPA) applications are widely used to process the ever increasing amounts of data streams produced by highly geographical distributed data sources such as fixed and mobile IoT devices in order to extract valuable information in a timely manner for real-time actuation. To efficiently handle this ever increasing amount of data streams, the emerging Edge/Fog computing paradigms is used as the middle-tier between the Cloud and the IoT devices to process data streams closer to their sources and to reduce the network resource usage and network delay to reach the Cloud. In this work, we account for the fact that both network resources and computational resources can be limited and shareable among multiple DSPA applications in the Edge-Fog-Cloud architecture, hence it is necessary to ensure their efficient usage. In this respect, we propose a resource-aware and time-efficient heuristic called SOO that identifies a good DSPA operator placement on the Edge-Fog-Cloud architecture towards optimizing the trade-off between the computational and network resource usage. Via thorough simulation experiments, we show that the solution provided by SOO is very close to the optimal one while the execution time is considerably reduced.

7.3 Timed protocol analysis of interconnected mobile IoT devices

Participants: Georgios Bouloukakis (Télécom SudParis), Nikolaos Georgantas (MiMove), Ajay Kattepur (Ericsson), Valerie Issarny (MiMove)

With the emergence of the Internet of Things (IoT), application developers can rely on a variety of protocols and Application Programming Interfaces (APIs) to support data exchange between IoT devices. However, this may result in highly heterogeneous IoT interactions in terms of both functional and non-functional semantics. To map between heterogeneous functional semantics, middleware connectors can be utilized to interconnect IoT devices via bridging mechanisms. In this work, we make use of the Data eXchange (DeX) connector model that enables interoperability among heterogeneous IoT devices. DeX interactions, including synchronous, asynchronous and streaming, rely on generic post and get primitives to represent IoT device behaviors with varying space/time coupling. Nevertheless, non-functional time semantics of IoT interactions such as data availability/validity, intermittent connectivity and application processing time, can severely affect response times and success rates of DeX interactions. We introduce timing parameters for time semantics to enhance the DeX API. The new DeX API enables the mapping of both functional and time semantics of DeX interactions. By precisely studying these timing parameters using timed automata models, we verify conditions for successful interactions with DeX connectors. Furthermore, we statistically analyze through simulations the effect of varying timing parameters to ensure higher probabilities of successful interactions. Simulation experiments are compared with experiments run on the DeX Mediators (DeXM) framework to evaluate the accuracy of the results. This work can provide application developers with precise design time information when setting these timing parameters in order to ensure accurate runtime behavior.

7.4 Performance analysis of Internet of Things interactions via simulation-based queueing models

Participants: Georgios Bouloukakis (Télécom SudParis), Ioannis Moscholios (University of Peloponnese), Nikolaos Georgantas (MiMove), Valerie Issarny (MiMove)

Numerous middleware application programming interfaces (APIs) and protocols were introduced in the literature in order to facilitate the application development of the Internet of Things (IoT). Such applications are built on reliable or even unreliable protocols that may implement different quality-of-service (QoS) delivery modes. The exploitation of these protocols, APIs and QoS modes, can satisfy QoS requirements in critical IoT applications (e.g., emergency response operations). To study QoS in IoT applications, it is essential to leverage a performance analysis methodology. Queueing-network models offer a modeling and analysis framework that can be adopted for the IoT interactions of QoS representation through either analytical or simulation models. In this work, various types of queueing models are presented that can be used for the representation of various QoS settings of IoT interactions. In particular, we propose queueing models to represent message-drop probabilities, intermittent mobile connectivity, message availability or validity, the prioritization of important information, and the processing or transmission of messages. Our simulation models demonstrate the significant effect on delivery success rates and response times when QoS settings are varied.

7.5 Consent-driven data use in crowdsensing platforms: When data reuse meets privacy-preservation

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

Crowdsensing is an essential element of the IoT; it allows gathering massive data across time and space to feed our environmental knowledge, and to link such knowledge to user behavior. However, there are major obstacles to crowdsensing, including the preservation of privacy. The consideration of privacy 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 introduce 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. We address this challenge with a clustering algorithm sensitive to the data distribution, which is shown to optimize data reuse and utility using a dataset from a deployed crowdsensing application.

7.6 PrioDeX: a Data Exchange middleware for efficient event prioritization in SDN-based IoT systems

Participants: Georgios Bouloukakis (Télécom SudParis), Kyle Benson (Real-Time Innovations), Luca Scalzotto (Injenia), Paolo Bellavista (University of Bologna), Casey Grant (National Fire Protection Association), Valérie Issarny (MiMove), Sharad Mehrotra (University of California, Irvine), Ioannis Moscholios (University of Peloponnese), Nalini Venkatasubramanian (University of California, Irvine)

Real-time event detection and targeted decision making for emerging mission-critical applications require systems that extract and process relevant data from IoT sources in smart spaces. Oftentimes, this data is heterogeneous in size, relevance, and urgency, which creates a challenge when considering that different groups of stakeholders (e.g., first responders, medical staff, government officials, etc) require such data to be delivered in a reliable and timely manner. Furthermore, in mission-critical settings, networks can become constrained due to lossy channels and failed components, which ultimately add to the complexity of the problem. In this work, we propose PrioDeX, a cross-layer middleware system that enables timely and reliable delivery of mission-critical data from IoT sources to relevant consumers through the prioritization of messages. It integrates parameters at the application, network, and middleware layers into a data exchange service that accurately estimates end-to-end performance metrics through a queueing analytical model. PrioDeX proposes novel algorithms that utilize the results of this analysis to tune data exchange configurations (event priorities and dropping policies), which is necessary for satisfying situational awareness requirements and resource constraints. PrioDeX leverages Software-Defined Networking (SDN) methodologies to enforce these configurations in the IoT network infrastructure. We evaluate our approach using both simulated and prototype-based experiments in a smart building fire response scenario. Our application-aware prioritization algorithm improves the value of exchanged information by 36% when compared with no prioritization; the addition of our network-aware drop rate policies improves this performance by 42% over priorities only and by 94% over no prioritization.

7.7 Traffic Refinery: Cost-Aware Traffic Representation for Machine Learning in Networks

Participants: Francesco Bronzino (Université Savoie Mont Blanc), Paul Schmitt (Princeton), Sara Ayoubi (Nokia Bell Labs), Hyojoon Kim (Princeton), Renata Teixeira (MiMove), Nick Feamster (University of Chicago)

Ever more frequently network management tasks apply machine learning on network traffic. Both the accuracy of a machine learning model and its effectiveness in practice ultimately depend on the representation of raw network traffic as features. Often, the representation of the traffic is as important as the choice of the model itself; furthermore, the features that the model relies on will ultimately determine where (and even whether) the model can be deployed in practice. This work develops a new framework and system that enables a joint evaluation of both the conventional notions of machine learning performance (e.g., model accuracy) and the systems-level costs of different representations of network traffic. We highlight these two dimensions for a practical network management task, video streaming quality inference, and show that the appropriate operating point for these two dimensions depends on the deployment scenario. We demonstrate the benefit of exploring a range of representations of network traffic and present Traffic Refinery, a proof-of-concept reference implementation that both monitors network traffic at 10 Gbps and transforms the traffic in real time to produce a variety of feature representations for machine learning models. Traffic Refinery both highlights this design space and makes it possible for network operators to easily explore different representations for learning, balancing systems costs related to feature extraction and model training against the resulting model performance.

7.8 Leveraging Website Popularity Differences to Identify Performance Anomalies

Participants: Giulio Grassi (MiMove), Renata Teixeira (MiMove), Chadi Barakat (Inria DIANA), Mark Crovella (Boston University)

Web performance anomalies (e.g. time periods when metrics like page load time are abnormally high) have significant impact on user experience and revenues of web service providers. Existing methods to automatically detect web performance anomalies focus on popular websites (e.g. with tens of thousands of visits per minute). Across a wider diversity of websites, however, the number of visits per hour varies enormously, and some sites will only have few visits per hour. Low rates of visits create measurement gaps and noise that prevent the use of existing methods. We develop WMF, a web performance anomaly detection method applicable across a range of websites with highly variable measurement volume. To demonstrate our method, we leverage data from a website monitoring company, which allows us to leverage cross-site measurements. WMF uses matrix factorization to mine patterns that emerge from a subset of the websites to "fill in" missing data on other websites. Our validation using both a controlled website and synthetic anomalies shows that WMF's F1-score is more than double that of the state-of-the-art method. We then apply WMF to three months of web performance measurements to shed light on performance anomalies across a variety of 125 small to medium websites.

7.9 LZR: Identifying Unexpected Internet Services

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

Internet-wide scanning is a commonly used research technique that has helped uncover real-world attacks, find cryptographic weaknesses, and understand both operator and miscreant behavior. Studies that employ scanning have largely assumed that services are hosted on their IANA-assigned ports, overlooking the study of services on unusual ports. In this work, we investigate where Internet services are deployed in practice and evaluate the security posture of services on unexpected ports. We show protocol deployment is more diffuse than previously believed and that protocols run on many additional ports beyond their primary IANA-assigned port. For example, only 3% of HTTP and 6% of TLS services run on ports 80 and 443, respectively. Services on non-standard ports are more likely to be insecure, which results in studies dramatically underestimating the security posture of Internet hosts. Building on our observations, we introduce LZR ("Laser"), a system that identifies 99% of identifiable unexpected services in five handshakes and dramatically reduces the time needed to perform application-layer scans on ports with few responsive expected services (e.g., 5500% speedup on 27017/MongoDB). We conclude with recommendations for future studies.

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), 2017-2021.

- "Application Performance Bottleneck Detection", Comcast Gift to Renata Teixeira, 2018-2021.

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 Venkatasubramanian)
- Inria MiMove (Valérie Issarny)

Duration: 2018 - 2022

URL: http://mines-mimove.inria.fr/

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 components 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 development 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 National initiatives

ANR BottleNet

Title: Understanding and Diagnosing End-to-end Communication Bottlenecks of the Internet

Partner Institutions:

- Orange
- LORIA
- Inria
- Télécom SudParis
- U. Lille 1

Duration: 2016 - 2021

Additional info: Bottlenet is about understanding and diagnosing end-to-end communication bottlenecks of the Internet. In particular, Bottlenet targets solutions that can be installed on users' devices (computers, tablets, phones, boxes or TVs) to capture users' perception in addition to the QoS measurements, in order to infer personalized models of their quality of experience. Beyond these measurements, Bottlenet intends to offer users a diagnostic service for their Internet connection based on the large-scale correlation of a multitude of measurements taken under different conditions in order to locate the source of a poor quality of experience. MIMOVE contribution focuses on an application-agnostic approach to QoE evaluation, where QoE is inferred only from information available from the observation of application traffic.

BPI – France Relance – 5G Events Labs

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: 6th ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI), CPS-IOT week, Virtual, May 2021; 2021 IEEE International Conference on Services Computing (SCC) as part of 2021 IEEE World Congress on Services (SERVICES), Virtual, September 2021; 8th IEEE International Conference on Smart Computing (SMARTCOMP 2022), Helsinki, Finland, June 2022.

Member of the organizing committees

- R. Teixeira, member of the selection committee of the Heidelberg Laureate Forum 2019–2021.

10.1.2 Scientific events: selection

Chair of conference program committees

- V. Issarny, TPC co-chair of: 7th International Conference on Smart Computing (SMARTCOMP), Virtual, June 2021; 18th Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), 2023.

- R. Teixeira, TPC co-chair of 18th USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2021.

Member of the conference program committees

- V. Issarny, member of the TPC of the following international conferences: IEEE ICDCS'21&'22, ACM/IEEE ICSE-SEIS'21, The Web Conference'21 (Senior PC), ACM ESEC/FSE'21, ACM/IFIP Middleware'21, IEEE CIC'21.
- N. Georgantas, member of the TPC of the following international conferences: ACM SAC'21&'22, SOSE'21, The Web Conference'21&'22, WETICE'21.
- N. Georgantas, member of the TPC of the following international workshop: SERENE'21.

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) - Co-chair of the Special Track for the COVID-19 and Future Pandemics [06/20–09/21]; IEEE Transactions on Software Engineering (TSE).
- N. Georgantas, member of the following editorial boards: MDPI Sensors - Internet of Things, International Journal of Ambient Computing and Intelligence (IJACI).

10.1.4 Leadership within the scientific community

- V. Issarny, Elected Chair of ACM Europe Council, 2021-2023.
- V. Issarny, Elected Secretary & Treasurer of ACM Europe Council, 2019-2021.

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).
- V. Issarny, member of Inria CR (junior researcher) and DR2 (Research Director position) selection committees, 2021.
- N. Georgantas, reviewer for the Leverhulme Trust Grant 2021 call for project proposals, reviewer for the Emergence 2021 call for project proposals, Sorbonne University Alliance.
- N. Georgantas, President of the EDITE Doctoral School selection committee "Communications, Networks and Systems" for PhD fellowships, 2021.

10.2 Teaching - Supervision - Juries

10.2.1 Supervision

- PhDs in progress:
  - Patient Ntumba (from August 2018): "Dynamic management of IoT data stream analytics in the edge-fog-cloud continuum", Sorbonne University, N. Georgantas and Vassilis Christophides (ENSEA).
  - 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.

10.2.2 Juries

- V. Issarny, member of the thesis committees of Claudia-Lavinia Ignat (HDR, University of Lorraine), Ilyas Toumlilt (PhD, Sorbonne University).
- N. Georgantas, member of the PhD mid-term committee of Benoit Martin (Sorbonne University), Laurent Prosperi (Sorbonne University), Razanne Abu-Aisheh (Sorbonne University).

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] G. Bouloukakis, K. Benson, L. Scalzotto, P. Bellavista, C. Grant, V. Issarny, S. Mehrotra, I. Moscholios and N. Venkatasubramanian. 'PrioDeX: a Data Exchange middleware for efficient event prioritization in SDN-based IoT systems'. In: *ACM Transactions on Internet of Things* 2.3 (July 2021), 1–32, Article No.: 19. DOI: 10.1145/3456301. URL: https://hal.archives-ouvertes.fr/hal-03171358.

[13] G. Bouloukakis, N. Georgantas, A. Kattepur and V. Issarny. 'Timed protocol analysis of interconnected mobile IoT devices'. In: *Journal of Internet Services and Applications* 12 (Dec. 2021), 12:1–12:31. DOI: 10.1186/s13174-021-00143-w. URL: https://hal.inria.fr/hal-03465847.

[14] G. Bouloukakis, I. Moscholios, N. Georgantas and V. Issarny. 'Performance analysis of Internet of Things interactions via simulation-based queueing models'. In: *Future internet* 13.4 (Apr. 2021), 87:1–87:13. DOI: 10.3390/fi13040087. URL: https://hal.archives-ouvertes.fr/hal-03184939.

[15] F. Bronzino, P. Schmitt, S. Ayoubi, H. Kim, R. Teixeira and N. Feamster. 'Traffic Refinery: Cost-Aware Data Representation for Machine Learning on Network Traffic'. In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 5.3 (14th Dec. 2021), pp. 1–24. DOI: 10.1145/3491052. URL: https://hal.archives-ouvertes.fr/hal-03489444.

International peer-reviewed conferences

[16] M. Brahem, G. Scerri, N. Anciaux and V. Issarny. 'Consent-driven data use in crowdsensing platforms: When data reuse meets privacy-preservation'. In: PerCom 2021 - IEEE International Conference on Pervasive Computing and Communications. Kassel / Virtual, Germany, 22nd Mar. 2021. URL: https://hal.inria.fr/hal-03097047.

[17] 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, 10th May 2021. DOI: 10.1109/INFOCOM42981.2021.9488832. URL: https://hal.inria.fr/hal-03109717.

[18] L. Izhikevich, R. Teixeira and Z. Durumeric. 'LZR: Identifying Unexpected Internet Services'. In: USENIX Security 2021 - 30th USENIX Security Symposium. Vancouver / Virtual, Canada, 11th Aug. 2021. URL: https://hal.inria.fr/hal-03143737.

[19] P. Ntumba, N. Georgantas and V. CHRISTOPHIDES. 'Efficient Scheduling of Streaming Operators for IoT Edge Analytics'. In: FMEC 2021 - Sixth International Conference on Fog and Mobile Edge Computing. Gandia, Spain, 6th Dec. 2021. URL: https://hal.inria.fr/hal-03413549.

[20] P. Ntumba, N. Georgantas and V. Christophides. 'Scheduling Continuous Operators for IoT Edge Analytics'. In: EdgeSys '21 - 4th International Workshop on Edge Systems, Analytics and Networking colocated with EuroSys'21. Online United Kingdom, United Kingdom: ACM, 26th Apr. 2021, pp. 55–60. DOI: 10.1145/3434770.3459738. URL: https://hal.inria.fr/hal-03208518.

Reports & preprints

[21] F. Bronzino, P. Schmitt, S. Ayoubi, H. Kim, R. Teixeira and N. Feamster. *Traffic Refinery: Cost-Aware Traffic Representation for Machine Learning in Networks*. 28th Jan. 2021. URL: https://hal.inria.fr/hal-03143736.