Journée des doctorants du département 3 (Réseaux, Systèmes et Services)
2017 PhD day of Department 3 (Networks, Systems and Services)

07/12/2017, room C005, LORIA

Program schedule

9:00 – 9:20 Opening session

9:00 – 9:05: Welcome (Ye-Qiong Song)

9:05 – 9:20: Explorer une thématique scientifique avec ISTEX (Jacques Ducloy)

9:20 – 10:35 session 1: Modeling and experimental methods

Abdulqawi Saif *(20' + 5' Q&A)*
Experimental methodology for evaluation of Big Data systems

Thomas Paris *(20' + 5' Q&A)*
Complex system modeling by composition

Giulia De Santis *(20' + 5' Q&A)*
Internet-Wide Scanners Classification using Gaussian Mixture and Hidden Markov Models

---

10:35 – 10:50 coffee break

10:50 – 12:30: session 2: large-scale systems

Abir Ismaili-Alaoui *(20' + 5' Q&A)*
Improving Business Process Management

Victorien Elvinger *(20' + 5' Q&A)*
Secured Convergence in Peer-to-Peer Collaborative Systems

Quentin Laporte-Chabasse *(20' + 5' Q&A)*
Toward a federated approach for web peer-to-peer collaborative systems

Maxime Compastié *(20' + 5' Q&A)*
Software-defined security for distributed cloud

12:30 – 14:00 Buffet & Discussion

14:00 – 15:45 session 3: Optimization and performance

Meihui Gao *(20' + 5' Q&A)*
Optimization models and methods for Network Functions Virtualization (NFV) architectures

Marjan Bozorg *(20' + 5' Q&)*
Optimization of architecture of membrane process

Guillaume Rosinosky *(20' + 5' Q&)*
Resource allocation and scheduling methods for BPMaaS providers

Hoai-Le Nguyen *(20' + 5' Q&)*
Studying group performance and behavior in collaborative editing

Matthieu Nicolas *(5')*
Efficient (re)naming in Conflict-free Replicated Data Types (CRDTs)

15:45 – 16:00 coffee break

16:00 – 18:05 session 4: Security

Nicolas Schnepf *(20' + 5' Q&A)*
Orchestration and verification of security functions for smart environments

Hoang Long Nguyen *(20' + 5' Q&A)*
End to End encrypted system for peer to peer collaborative editing

Amina Ahmed Nacer *(20' + 5' Q&A)*
Contribution to the secure deployment of business processes in the cloud

Xavier Marchal *(20' + 5' Q&A)*
Secure operation of virtualized Named Data Networks

Daishi Kondo *(20' + 5' Q&A)*
Risk Analysis of Information-Leakage through Interest Packets in NDN

Appendix: Abstracts of the presentations

(in order of the program schedule)

Openning session

Jacques Ducloy :

Explorer une thématique scientifique avec ISTEX

Résumé : Dans le cadre du projet ISTEX (financé par les Investissements d'Avenir), près de 20 millions d'articles en texte intégral sont mis à la disposition de l'ESR. Une application particulièrement intéressante pour les doctorants est la possibilité d'exploiter de vastes corpus (plusieurs milliers de documents) pour y explorer un domaine scientifique. Ceci permet d'élargir considérablement la bibliographie d'une thématique scientifique. Nous présenterons la plateforme LorExplor qui est développée dans cette perspective. Elle sera mise (avec assistance) à la disposition des doctorants intéressés.

Session 1

Abdulqawi Saif

Experimental methodology for evaluation of Big Data systems

Abstract: Big Data systems and applications are taking places over traditional systems in many domains. This trend influences the market to build a lot of different-architecture systems to deal with all Big Data life-cycle phases i.e. generating, analyzing and storing massive data. However, experimenting on this kind of systems is not yet mature since we are still using the existed experimental methods that are not targeting the complexity of Big Data systems. Thus, this thesis aims at proposing effective experimenting methods which address the needs of Big Data systems as well as the needs to efficiently using the resources of evaluation environment. Several contributions are proposed so far in the context of this thesis. Firstly, we addressed the problem of wasting testbed resources while performing Big Data experiments. Experimenters try to orchestrate all the experimental factors at once which leads to increase the number of experiments hence to use more resources. Thus, we proposed to use a statistical methodology to reduce the number of experiments in any experimental context while still reporting an acceptable level or results. This method is tested against NFS protocol in a high latency WAN environment [1]. Then, we proposed IOscope [2] to uncover the I/O access patterns of in-production databases. We were motivated to propose IOscope since analyzing the pattern of I/O accesses is difficult, as they are performed by the deeper layers of applications, and interact directly with the kernel interfaces. Furthermore, existing tools that attempt to provide insight into I/O access generally have a high overhead which excludes them from being used for in-production systems. Currently, we are working on proposing an automated methodology to monitor experiments on testbeds, exploiting the monitoring results to produce figures that go with scientific publication process. Experimenters used to rely on infrastructure monitoring tools to monitor their experiments which requires additional efforts and manual actions.

[1] Saif, A., & Nussbaum, L. (2016, July). Performance Evaluation of NFS over a Wide Area Network. In COMPAS-Conférence d'informatique en Parallélisme, Architecture et Système.

[2] SUBMITTED BUT NOT YET ACCEPTED: Saif, A., Nussbaum, L., & Song, Y.Q. IOscope: Uncovering the I/O Access Patterns of In-Production Databases. In EDBT (2018).

Thomas Paris

Complex system modeling by composition

Abstract: Complex systems are systems composed of many heterogeneous interacting entities. For example, a smart-grid is a complex system which involves entities from at least three domains (power grid, telecommunication and information system). Multi-modeling (modeling each part of the system individually) and co-simulation (simulating each part separately and ensuring synchronization and data-exchange between simulators) are promising approaches to handle complexity but request to handle heterogeneities (domains, formalisms...).

MECSYCO (Multi-agent Environment for Complex System CO-simulation) is a co-simulation middleware with an integrative approach (reuse of legacy, pre-existing or even future systems) based on formal and software wrapping. It provides means to handle het- erogeneities and to ensure modularity.

One of the challenges is to enable this integrative approach in a multi-modeling and co-simulation process while providing together the handling of heterogeneities, the modularity and the closure under coupling property (composition of models allowing hierarchical construct).

The purpose of my work is then to enhance MECSYCO with a multi-modeling activity support and the closure under coupling property while maintaining its advantages. To meet this goal, we define a multi-modeling and co-simulation process based on description files and we explore the use of Domain Specific Languages to support each step of this process. This work opens new perspectives concerning multi-model checking, co-simulation deployment...

Keywords: Modeling & Simulation, Composition, Complex System

Giulia De Santis

Internet-Wide Scanners Classification using Gaussian Mixture and Hidden Markov Models

Abstract: Internet-wide scanners are heavily used for malicious purposes: exploitation of vulnerabilities, control over the host, gather of information about the system. Since they are often used during the reconnaissance phase of Advanced Persistent Threats (APTs), fingerprinting them allows security experts to detect firstly the faced scanning technique, and secondly that an attack is probably ongoing. Our work models, from the scanned system point of view, three features of the network scanning activities: intensity and spatial and temporal movements, which are related to the number of scanned IP addresses in a given time window, and the difference of successive scanned IP addresses and timestamps, respectively. Based on real logs of incoming IP packets collected from a darknet, Hidden Markov Models (HMMs) are used to model scanning techniques. The ones built on spatial and temporal movements are then used to assess what scanning technique is operating. The proposed methodology, using only one of the latter two aforementioned features of the scanning technique, is able to fingerprint what network scanner originated the perceived darknet traffic.

Session 2

Abir Ismaili-Alaoui

Improving Business Process Management

Abstract: Business Process Management (BPM) is concerned with continuously enhancing business processes by adapting a systematic approach that enables

companies to evolve the performance of their existing business processes and achieve their business goals.
By studying several business process models we realize that business processes are in general considered as blind and stateless, which mean that in each business process execution we do not take into consideration the results from last process instances.
The main objective of our research is to exploit the data generated from previous instances in order to enhance business processes in several aspects, such as resource allocation and scheduling. In general, business Processes are different from scientific workflows as they may contain automatic tasks and non automatic tasks, so managing resources effectively in business processes depends on the type of those resources (Human or machine).
In this optic, our work consists on scheduling business process in- stances based on the priority of the events that launch those instances, using machine learning to analyze data generated from previous business process execution. This step ensure the assignment of the most critical business process instance tasks, to a qualified human resource while minimizing execution costs of this type of tasks for the company.

Victorien Elvinger *(20' + 5' Q&A)*
Secured Convergence in Peer-to-Peer Collaborative Systems
Abstract: In peer-to-peer collaborative systems, participants modify their own copy of co-authored data. Those copies continuously diverge then converge according to the exchange between participants. Convergence is a liveness property of collaborative systems. A Byzantine adversary may conceal divergence from honest participants. Authenticated logs make misbehavior of malicious participants evident and thus preserve convergence of the copies owned by honest participants. In dynamic groups, all participants have to store the entire log of contributions in order to invite newcomers. Indeed, newcomers must retrieve and play this log to get the current state of the data. This implies large storage, communication, and computation overheads.
We propose a protocol that exchanges and authenticates an untrusted snapshot of the data using a trusted and pruned version of the log of contributions. In order to prune the log, we extend the concept of causal stability to dynamic groups and adapt View-Fork-Join-Causal (VFJC) to perform consistency verification. Our protocol reduces both the storage footprint of the log and the joining cost for a newcomer.

Quentin Laporte-Chabasse
Toward a federated approach for web peer-to-peer collaborative systems
Abstract : Real-time editing collaborative applications aim to provide a reliable way to share and edit content for remote users. A major part of those kinds of applications are web applications based on centralized architecture. This means that an authority ensure communication and persistence of data provided by collaborators. This paradigm raises issues regarding the scalability of the application and user's privacy.
In this context, a new kind of web collaborative application relying on peer-to-peer communication paradigm arises. Thus, collaborators can directly collaborate and share without passing through any authority. Scalability and privacy are therefore solved by this way.
On a different scale, we can consider several organizations who possess their own peer-to-peer collaborative network. In some instances, those organizations have to collaborate among themselves to achieve a common goal. On the one hand, it is

necessary to ensure availability of the infrastructure for all implied organizations and, on the other hand, it is crucial to preserve the sovereignty of each organization.
We propose a first approach allowing to federate organizations. This approach is divided in three steps focusing respectively on the federation of joining infrastructure, roles and policies user centered, mechanisms to ensure availability against churn. This approach is partially drawn from DOSNs (Distributed Social Network) which deal with similar social and technical issues.
Key words : *collaborative editing, federation, peer-to-peer infrastructure*

Maxime Compastié
Software-defined security for distributed cloud
Abstract: While cloud computing provides new facilities for building elaborated services, the distribution of related resources across multiple infrastructures (multi-cloud) and several tenants (multi-tenancy) challenges their management, in particular from a security perspective. Moreover, the cloud resource heterogeneity induces a substantial overhead in the selection, allocation and configuration of the adequate security mechanisms. In that context, the software-defined security (SDSec) paradigm tackles management issues by uncoupling security policy-based management from the enforcement through programmable security interfaces.
We propose and justify a software-defined security strategy to propose a cloud-wide enforcement accounting for multi-tenancy and multi-cloud constraints.
We take advantage of unikernel system architecture properties in the virtualization context to support this enforcement and provide resources with low attack surface embedding finely crafted and adapted security mechanisms. These resources correspond to highly constrained configurations with the strict necessary for a given time-limited period.
We describe the management framework supporting this software-defined security strategy, formalizing the generation of unikernel images that are dynamically built to comply with security requirements over time. Through our implementation, and extensive experiments, we show that the cost induced by our security integration mechanisms is small while the gains in limiting the security exposure are high.
Keywords: Security Management, Cloud Infrastructures and Services, Software-defined Security, Resource Virtualization, Unikernel.

Session 3

Meihui Gao
Optimization models and methods for Network Functions Virtualization (NFV) architectures
Abstract: Due to the exponential growth of service demands, telecommunication networks are populated with a large and increasing variety of proprietary hardware appliances increasing the cost and the complexity of the network management. The NFV paradigm is proposed to overcome this problem, allowing dynamical allocation of Virtual Network Functions (VNFs). A key problem in NFV is the VNF chaining: given a network where some nodes are connected with computational servers and a set of demands asking for a sequence of VNFs, VNF instances need to be installed on servers and the demands must be routed in such a way that each demand accesses the requested VNFs. Different objectives can be considered, one of the most interesting is to reduce the cost.

From an optimization point of view, the problem can be modeled as the combination of a network design problem (for the demands routing) and a facility location problem (for the VNF location and server dimensioning). Both problems are widely studied in the literature, but their combination represent a new challenge. The goal of this thesis is to investigate the computational complexity and properties of the VNF chaining problem and to propose and compare different mathematical programming models. Furthermore, to develop different solution methods (exact and heuristics) to solve the problem.

Marjan Bozorg

Optimization of architecture of membrane process

Abstract: Membrane network for separating multicomponent gas mixtures is a very promising separation technology. To optimize the performances and the global costs of such processes, the development of accurate and reliable design strategy is essential. Finding an optimal design for a membrane system is a complicated task, because of many characteristics that must be taken into account.

The problem of the optimization of a membrane system is classified as a non-convex Mixed Integer Nonlinear Programming (MINLP) problem. In fact, non-linarites are induced by the single membrane system properties and combinatorial decisions are necessary to determine the design.

The objective of this thesis is to develop a methodology which aim is to design the optimal structure of membrane systems in order to simultaneously minimize operating and/or capital costs while guaranteeing separation performances of the system. This thesis define as a multidisciplinary project in collaboration between the reaction laboratory and process engineering (LRGP) of University of Lorraine, optimization team of LORIA and engineering faculty of University of Rome Tor Vergata

We decided to firstly investigate the non-convex continuous nature of the problem, and propose a global optimization algorithm to tackle this problem. Our global optimization strategy is now under investigation of different case studies.

Guillaume Rosinosky

Resource allocation and scheduling methods for BPMaaS providers

Abstract: With the generalization of cloud computing, providers can distribute their software as a service, and customers can benefit from it without investing in large and expensive infrastructures. However, without efficient resource allocation methods, operational cost can rise very quickly. This is also the case for BPM as a Service providers wanting to be able to manage elastically hundreds of customers for a sufficient quality of service. This is not an easy task regarding the different cloud providers price models, different resources types, and offers. In our work, we propose resource allocation and scheduling methods based on heuristics and integer linear programming for BPMaaS providers. Our approach is tenant-based instead of process-based as in many of the related work, and use BPM task throughput as a general estimation of needed computing power for tenants and as the capacity for cloud resources. We also include migrations between cloud resources number limitation as a constraint, as it can provoke quality of service breaks. Our approaches are offline and could also be used with other type of software. Current experiments concern the metrics enhancement, migrations impact on quality of service estimation, and tests of our algorithms coupled with an orchestrator.

Hoai-Le NGUYEN

Studying group performance and behavior in collaborative editing

Abstract: Collaborative editing (CE) allows a group of people to modify a shared document simultaneously (real-time) or non-simultaneously (asynchronous). It has gained in popularity with several free services and tools designed to support collaborative editing such as Google Drive, Wikipedia and Version control systems. Collaboration traces collected from these systems represent a valuable data for research. Our objective is to study group performance and behavior in CE by analyzing collaboration traces. The questions we address are: How often do conflicts happen in CE? How does automatically-resolved process support CE? How do users react to conflicts during CE? Answering these questions is important to ensure good performance and user experience. Each research question needs to be considered in both real-time and asynchronous context. Compared to asynchronous CE, in real-time CE, each collaborator receives immediately updates of other users and is therefore aware of the parts of the document being edited. However, when the number of collaborators increases, delays for seeing other users modifications are larger and conflicts still happen. Moreover, in real-time collaboration, performances are considered more important than in asynchronous collaboration. So far we analyzed the different effects of conflicts and user behaviors in asynchronous CE on collaboration traces that used Git.

Matthieu Nicolas

Efficient (re)naming in Conflict-free Replicated Data Types (CRDTs)

Abstract: In order to design large scale distributed systems, the literature and companies increasingly adopt the optimistic replication model known as eventual consistency to replicate data among nodes. This consistency model allows replicas to temporarily diverge to be able to ensure high availability. Each node owning a copy of the data can edit it without any kind of coordination with other nodes, before propagating the changes to others. A conflict resolution mechanism is however required to handle updates generated in parallel by different replicas.

An approach which gains in popularity since a few years proposes to define Conflict-free Replicated Data types (CRDTs). These data structures behave as traditional ones, like the *Set* or the *Sequence* data structures, but are designed for a distributed usage. Their specification ensures that concurrent changes are resolved deterministically and that replicas eventually converge after observing all updates.

To achieve convergence, CRDTs proposed in the literature mostly rely on identifiers to reference updated elements. To be globally unique, element identifiers often include the identifier of the node which generates them. But, since node identifiers grow as new nodes join the system, element identifiers have to grow proportionally. Furthermore, element identifiers have to comply to additional constraints according to the CRDT, which may result in the acceleration of their growth.

Hence, since the size of identifiers is not bounded, the size of metadata attached to each element increases over time. It thus exceeds more and more the size of data itself. This impedes the adoption of CRDTs since nodes have to broadcast and store metadata, causing the application's performances and efficiency to decrease over time.

The goal of this PhD is to address this issue by 1. proposing more efficient specifications of identifiers according to their set of constraints, 2. proposing mechanisms to rename identifiers to reduce their size.

Session 4

Nicolas Schnepf

Orchestration and verification of security functions for smart environments

Abstract: Security threats against smart environments are exponentially growing for several years due to lack of market preventive methods. A solution proposed by researchers consists in chaining security functions for dynamically protecting those devices: in particular, those chains would benefit of the programmability provided by software defined networks (SDN) for automating their deployment and their adjustment. Nevertheless, the multiplication and the complexity of such chains of security functions increase the risk of introducing misconfigurations in network policies: because of this complexity, the validation of such chains require the use of formal methods for guarantying their correctness before their deployment. The goal of this PhD is to design a framework for the orchestration and the verification of chains of security functions. In our previous work we already designed an approach for the validation of security policy called synaptic: this framework relies on formal methods for validating the correctness of SDN policies. Complementary to this work we proposed an approach for automatically profiling android applications in order to identify their security requirements. The remaining part of our work will consist in designing an approach for automatically generating or selecting chains of security functions corresponding to the applications running on a device.

Hoang Long Nguyen

End to End encrypted system for peer to peer collaborative editing

Abstract: My thesis focuses on designing an end-to-end encrypted system for peer-to-peer (P2P) collaborative editing among users from different enterprises that maintain their own security servers. The system should protect communication over network from eavesdroppers, including ISPs, system providers and even the enterprises themselves while being scalable and easy to use.

The challenges addressed in my thesis are: (1) establishing an autonomous, scalable key management system which prevents adversaries to perform Man-in-the-Middle attack by equivocating public keys of users and (2) designing a fine grain decentralized access control mechanism for P2P collaborative editing systems that can cope typical challenges in distributed system such as race condition and byzantine behaviors.

My presentation will focus on the first challenge related to the design of a transparent-log scheme to ensure the trustworthiness of enterprise key servers. We proposed an auditing mechanism for transparent-log system using public blockchain. Compare to related work, our proposal is easy to implement, inexpensive to operate and resilient to malicious clients.

Amina Ahmed Nacer

Contribution to the secure deployment of business processes in the cloud

Abstract: The constant development of technologies forces companies to be innovative in order to stay competitive. Therefore, companies are ready to outsource their business processes to the cloud to enjoy its benefits, but they are reluctant to expose their BP models which express the know-how of their companies.

In fact, preserving privacy of organizations is a huge challenge. It is becoming one of the main society concerns, especially when outsourcing business processes. And probably

that what is yet a problem in general, is even more exacerbated in the context of business processes (BP) because of the sensitive information they include about their organizations.
In this optic, our work consists on developing approaches allowing to securely deploy BP in the cloud while preserving the BP-Know-how and decisions strategy of companies.

Xavier Marchal
Secure operation of virtualized Named Data Networks
Abstract: A new trend has recently emerged in computer networking named Network Function Virtualization (NFV). NFV is defined by the European Telecommunications Standards Institute (ETSI) and this institute is in charge of the standardization of the NFV concept. NFV involves the implementation of network functions as software, named Virtualized Network Function (VNF), that can be run on a wide range of hardware (mainly on x86 based servers). It allows a lot of flexibility in the conception of physical network and is more affordable because the hardware is no more proprietary (mass production) and the VNFs can be mixed together thanks to the NFV standard (economic competition). NFV takes the same advantage of virtualization as traditional virtual machines and will help network operators to deploy more efficient and robust networks but also smoothly deploy new network functions or protocols that do not exist yet. New services can also emerge with NFV like network-as-a-service that can be provided by could operators.
Another long term research subject is to design new protocols that can handle more efficiently the current needs of computer communications. One of these is named Named Data Networking (NDN) that is part of Information Centric Network (ICN) family. NDN is seen as an alternative to traditional TCP/IP networks by directly addressing contents instead of hosts and its main use case is Internet. However, such a deployment is difficult to envisage on dedicated hardware. Many questions about the security, performance and interoperability of this new protocol need to be solved before it can happen. NFV could provide a new and preferred framework for addressing the problems slowing down the adoption of these new protocols by enabling reliable and secure use of VNFs implementing new network architectures such as NDN.
The thesis aims to design virtualized network functions that meet the operational requirements (security, performance, interoperability) required by the operators and must be ensured before deploying new solutions. In addition, the management and orchestration of NDN services deployed as VNFs will ensure that network services are properly deployed to achieve the desired quality of service while maintaining the security of the underlying infrastructure and hardware resource management.

Daishi Kondo *(20' + 5' Q&A)*
Risk Analysis of Information-Leakage through Interest Packets in NDN
Abstract: Information-leakage is one of the most important security issues in the current Internet. In Named-Data Networking (NDN), Interest names introduce novel vulnerabilities that can be exploited. By setting up a malware, Interest names can be used to encode critical information (steganography embedded) and to leak information out of the network by generating anomalous Interest traffic. This security threat based on Interest names does not exist in IP network, and it is essential to solve this issue to secure the NDN architecture. This paper performs risk analysis of information-leakage in NDN. We first describe vulnerabilities with Interest names and, as countermeasures, we propose a name-based filter using search engine information, and another filter

using one-class Support Vector Machine (SVM). We collected URLs from the data repository provided by Common Crawl and we evaluate the performances of our per-packet filters. We show that our filters can choke drastically the throughput of information-leakage, which makes it easier to detect anomalous Interest traffic. It is therefore possible to mitigate information-leakage in NDN network and it is a strong incentive for future deployment of this architecture at the Internet scale.