COST ACROSS Workshop on

Engineering and Control for Reliable Cloud Services

Co-located with ITC27, Ghent, Belgium, 11 September 2015

Abstracts of presentations and speakers' bios

Challenges in Developing an Efficient Cloud Management Framework (Keynote) – Eliot Salant
Cloud Computing has developed rapidly over the last ten years. World-wide spending on public and private cloud hosting is expected to pass the $32bn. mark this year as businesses switching to the Cloud typically can reduce their CAPEX and OPEX budgets by better than 30% due to the cloud's ability to better take advantage of economies of scale. Yet in actuality, in commercial data centres the utilization of resources still remain low, leaving a lot more room for cost savings and a reduced energy footprint. In this talk we will introduce the challenges in managing both the cloud infrastructure and the application more effectively to obtain better utilization of the cloud ecosystem, including the challenges involved in multi-tenancy issues in placement of an application, sizing an application, adaptation of the infrastructure to improve workload performance and monitoring analytics of the ecosystem.

Eliot Salant is a Research Staff Member at IBM Research - Haifa. He has been the Project Coordinator for four large, highly successful FP7 projects, including RESERVOIR, VISION Cloud and the currently running CloudWave in the area of Cloud Computing, and ENSURE in the area of long term digital preservation. He has been at IBM for over twenty years, and has held a wide range of managerial and lead technical roles. Eliot has a B.Eng in Mechanical Engineering from McGill University, a MSc. in Biomedical Engineering from The Technion - Israel Institute of Technology and a MSc. in Computer Science from Union College.

Energy Efficient Control of Virtual Machine Consolidation under Uncertain Input Parameters for the Cloud – Andreas Kassler
Reducing the energy consumption of datacenters and the Cloud is very important in order to lower CO2 footprint and operational cost of a Cloud operator. However, there is a tradeoff between energy consumption and perceived application performance. In order to save energy, Cloud operators need to consolidate as many Virtual Machines (VM) on the fewest possible physical servers. However, such consolidation policy may lead to violation of SLA if VMs suddenly run at their peak demand and the number of used servers was too small to cope with such overload. Such consolidation is facilitated in modern datacenters by using live VM migration techniques, which stress the network. As a consequence, it is important to find the right balance between the energy consumption of the physical servers, the placement of the VMs after the consolidation process and the number of migrations to perform. What makes such VM consolidation challenging is the service reliability and SLA requirements of VMs given the highly dynamic nature in modern datacenters, where many aspects have a high degree of uncertainty. In this work, we therefore propose a novel approach to the energy efficient VM

consolidation problem based on the theory of Robust Optimisation (RO). By applying this theory, we develop a mathematical model of the energy efficient VM consolidation problem as a robust Mixed Integer Linear Program and specify uncertainty bounds for resource requirements of the VMs and the rack server power model. Those bounds are inspired by a measurement campaign done in our universities datacenter observing CPU workload of 6 different VMs that implement various parts of our economic system and authentication infrastructure. We show that our model allows to tradeoff two important aspects for a datacenter operator: by taking higher risk aversion, our model will take into account more severe and unlikely deviations of the uncertain parameters, leading to higher protection but also higher energy consumption. Alternatively, if one wants to take a higher risk, the solution will offer less protection at lower energy consumption

Dr. Andreas Kassler received his MSc degree in Mathematics/Computer Science from Augsburg University, Germany in 1995 and his PhD degree in Computer Science from University of Ulm, Germany, in 2002. Currently, he is employed as Full Professor with the Department of Mathematics and Computer Science at Karlstad University in Sweden. Before joining Karlstad University, he was Assistant Professor at the School of Computer Engineering, Nanyang Technological University, Singapore, between 2003 and 2004. Dr. Kassler is (co-)author of more than 100 peer reviewed books, journal and conference articles. He served as a guest editor of a feature topic in EURASIP Wireless Communications and Networking Journal, and is on the editorial boards of several international journals. Dr. Andreas J. Kassler is a member of IEEE Computer Society and IEEE Communications Society.

Resource Pricing and Provisioning Strategies in Cloud Systems: A Stackelberg Game Approach – Francesco Lo Presti

In this talk, we consider several Software as a Service (SaaS) providers that offer services using the Cloud resources provided by an Infrastructure as a Service (IaaS) provider which adopts a pay-per-use scheme, comprising flat, on demand, and spot virtual machine instances. For this scenario, we study the virtual machine provisioning and spot pricing strategies: We assume that the SaaS providers want to maximize a suitable utility function which accounts for both the QoS delivered to their users and the associated cost; The IaaS provider, on the other hand, wants to maximize his revenue by determining the spot prices given the SaaS bids. We consider a two-stage provisioning scheme. In the first stage, the SaaS providers determine the optimal number of required flat and on demand instances. Then, in the second stage, the IaaS provider sells its unused capacity as spot instances for which the SaaS providers compete by submitting a bid. We study two different IaaS provider pricing strategies: the first assumes the IaaS provider sets a unique price; in the second, instead, the IaaS provider can set different prices for different users. We model the resulting problem as a Stackelberg game. In this class of games, one player (i.e., the leader, in our case the IaaS provider) moves first and commits his strategy to the remaining players (i.e., the followers, in this case the SaaS providers), that consider the action chosen by the leader before acting simultaneously to choose their own strategy. The adoption of a leader follower strategy allows us to study revenue-maximizing pricing scheme for the IaaS provider. We address the solution of the resulting Stackelberg game by solving a suitable Mathematical Program with Equilibrium Constraint (MPEC) problem. For each pricing scheme, we show the existence of the game equilibrium and provide the solution algorithms. Finally, we illustrate through numerical investigation the behavior of the proposed provisioning and pricing strategy under different workload and bidding configurations.

Francesco Lo Presti is an Associate Professor in Computer Science in the Department of Civil Engineering and Computer Science of the University of Rome “Tor Vergata”. He received the MSc in Electrical Engineering and the PhD in Computer Science in 1997 at the University of Rome "Tor Vergata". From 1998 to 2001, he held a post-doc position at the University of Massachusetts at

Amherst and the AT&T Research Labs. His research interests include performance modelling, analysis and optimization of computer networks. He has been involved in several European Projects in the ICT area. He has more than 50 publications in international conferences and journals. He has served as a program member of international conferences on networking and performance areas, and serves as reviewer for various international journals.

Price/Quality Integrated Modelling of Cloud-based Composite Services Performance – Ivan Ganchev

A generic, integrated Service Network Model (SNM) is proposed for a fully end-to-end composite services performance modelling. Two main types of composite service provision are considered: (i) with Quality of Service (QoS) guarantees, e.g. based on resource reservation, and (ii) without QoS guarantees, i.e. using resources, which have been dynamically released by other users demanding QoS guarantees. The proposed performance model, based on the teletraffic theory, utilizes mathematical equations and computer simulations, and considers the following three sets of parameters: **UB**: User behavior defining the demand for services and including parameters such as service requests frequency, offered set of service types, user persistence, etc.; **TCH**: Technical characteristics of the service network including type and capacity of resources (transmitting-, switching-, computing-, storage resources, etc.); **QoS**: Service-related QoS parameters, such as service response time, availability, security, reliability, support level, reputation/user ranking etc., and network-related QoS parameters, such as network losses, delay, throughput, etc. The model allows solving problems even if some parameters of any one of the sets above are unknown, provided all parameters of the other two sets are known. This could be highly useful for service network analysis, control, performance optimization, and strategic management including dimensioning/re dimensioning, QoS forecasting, recommendations for influencing the user behavior, etc. In addition, an econometric extension of the model is discussed for ensuring the service network will work efficiently from both business and technical perspectives' point of view, based on the following parameter sets and dependences: **SPC**: Service prime cost as a function of UB, TCH, SEI, and service network performance (QoS); **QoE**: Quality of Experience as a function of QoS and SP; **SP**: Service price as a function of SPC, UB, and SEI; **UB**: User behavior as a function of SEI and SP/QoE ratio; **SEI**: Socio-economic environment. The current research results and directions, following the described approach, are discussed.

Ivan Ganchev <<< bio >>>

Modelling Cloud Federations – Wojciech Burakowski

Cloud federation concept is regarded as a promising solution to essentially extend the capabilities of alone operating clouds. These extensions correspond to such characteristics as greater volume of available resources (data bases, compute machines etc.) and, as a consequence, in offering wider range of cloud services. However, in order to make a cloud federation to become efficient, we need in that case to design more complicated management system comparing to this, which we have for alone operating clouds. This happens due to two important factors that have great impact on cloud federation behavior, which are: (1) control of network connecting clouds, and (2) management of all resources that in fact belong to different clouds. On the other hand, to understand which mechanisms and algorithms are suitable to manage cloud federation, we need to define appropriate model for cloud federation. It appears that such model becomes quite complex and it should take into account a lot of components and it should correspond to different time scales. In the paper we present a complete model allowing to perform studies on cloud federation. It includes such issues as e.g.: (1) strategies for handling user requests (requests for service provisioning and requests for service execution), (2) network aspects as network dimensioning and network control, (3) task allocation strategies for service provision and service execution, and (4) rules for resource sharing. The model is

suitable for performing simulation studies as well as for developing analytical approaches. The paper also discusses some possibilities to perform practical experiments corresponding to cloud federation in European testbeds, which were developed by projects from 7FR and Horizon 2020. Finally, exemplary simulation results will be presented and discussed.

Wojciech Burakowski takes full professor position at the Institute of Telecommunications, Warsaw University of Technology since 2006. He is the head of the Internet Architectures and Application Division. He is author or co-author of about 200 research publications in national and international conference proceedings, journals and books. He is TPC member of many international and national conferences dealing with computer systems and networking. His main research areas includes multi-service networks, architectures and protocols, performance evaluation and designing prototypes. Since 1990 he actively participates in European projects as COST 242, 242, 257 and 279, all related to the developing methods for performance evaluation of different network technologies, as ATM, IP, wireless networks etc. He participated in the EU 5FR, 6FR and 7FR projects targeted for building prototypes of QoS networks for single and multi-domains. In 2010-2013 he was leading the strategic national research project entitled "Future Internet Engineering", gathering more than 120 researchers from 9 top research centers and technical universities in Poland. The results of the projects were presented in many international events, among them on CEBIT 2013. Currently he is involved in national project "PL-LAB2020" aimed at building national testbed and COST IC1304 "Autonomous Control for a Reliable Internet of Services (ACROSS)".

NFV and Telco-Clouds: Introduction and Challenges – Erwin Six

Over the past years, cloud technologies have drastically changed the way the telecommunication industry thinks about their network, their services and their operations. In this presentation various technology aspects of Network Function Virtualization (NFV) and Software Defined Networking (SDN) will be explained together with practical use-cases on how all of this will be deployed by operators in the near future. Beside the pure technical aspects, it will also bring insights in why certain technologies where crucial in disrupting fundamental business needs in the industry, and which challenges are ahead of us.

Erwin Six is leading the Cloud/NFV research team in Bell Labs Belgium. He received a master's degree in Electrical Engineering from the Ghent University in 2001. After university, he joined the Research & Innovation department of Alcatel, which became part of Bell Labs in 2006. Within both research organizations he has been in different roles – researcher, technical manager, project leader and department head. His research was focused on a number of technical areas within the domain of telecommunication: Passive Optical Networks, Ethernet & IP Access Networks, Home network Management, Service enabled network architectures, video technologies and cloud. His innovations resulted in the creation of various new Alcatel-Lucent products: e.g. 7342 GPON, 7302 Intelligent Services Access Manager (ISAM), 5580 Home Network Manager and the 5910 Video Service Appliance and Immersive Communication. Besides this, his research had impact on different standardization activities (FSAN GPON, DSLForum,…). Erwin Six is (co-)author of around 15 papers and more than 30 patents. In 2006 he won the Alcatel Patent award.

Cloud Services in Community Networks and by Citizen Participation – Felix Freitag

An operational geographically distributed and heterogeneous cloud infrastructure with services and applications deployed in Guifi.net, the probably largest community network worldwide with more than 25000 nodes, is presented. It is particular case of a community cloud, developed according to the specific needs and conditions of community networks. We describe the concept of this community

cloud, explain our technical choices for building it, and our experience with the deployment of this cloud. In a second part we discuss several research challenges that we found and that still need to be solved for building such a community cloud, and for making it sustainable.

Felix Freitag <<< bio >>>

QoE for Cloud Services – Martín Varela

Quality of Experience (QoE) has become an important concept in service delivery, as a good understanding of it allows for a good understanding of how happy customers are with a service, and this in turn directly affects the bottom line of the service provider. In the context of Cloud services, where competition is oftentimes based on cost, and performance guarantees are lackluster, QoE models can be applied to the provider's (be them IaaS, PaaS, or SaaS) advantage, with applications ranging from resource management to pricing and service differentiation. Currently, however, there are several challenges to overcome before QoE can be fully exploited by Cloud service providers. First and foremost, while QoE is very well understood for media services, it is significantly harder to come up with good QoE models for other types of services. Secondly, being able to utilize QoE estimations to their full potential often requires cooperation between multiple parties, whose goals are not necessarily aligned. Finally, the way that services are marketed needs to undergo a change of mindset, moving from a focus on price, to a focus on quality and utility. In this talk we will briefly cover the basics of QoE, as well as the motivation for considering it as a business tool for Cloud service providers. We will also cover some of the challenges related to QoE in the Cloud, and some ideas on how they could be overcome.

Martín Varela received his PhD and MSc from the University of Rennes 1 (Rennes, France), in 2005 and 2002 respectively. He has been an ERCIM fellow, and spent time at SICS and VTT, where he is currently a Principal Scientist and leads the current work on QoE. He has recently been a guest Senior Researcher at the Cooperative Systems Group at the University of Vienna. Dr. Varela was a Finnish management committee member for the recently ended COST Action IC1003 Qualinet, and is a substitute management committee member for COST Action IC1304 ACROSS. He currently co-chairs the IEEE MMTC QoE Interest Group. His research interests lie in the QoE domain, with a particular focus on real-time QoE models and QoE management and their applications, both in the technical and business domains.

Towards Multi-layer Performance Management in Cloud Networking and NFV Infrastructures – Kurt Tutschku

Cloud Networking (CN) offers an appealing novelty for Cloud Computing (CC) customers. They can do a one-stop-shopping for complex Cloud services that are enhanced with clever network service features. These features result from a tight integration of cloud applications and smart networks, e.g. using application-specific Virtual Private Networks or proxies for adaptive HTTP-based streaming. NFV (Network Function Virtualization) infrastructures, which are quite similar in their technologies and architecture with CN systems, are currently revolutionising telecommunication networks due to the introduction of virtualisation and data centre technologies to the control and operation of public communication networks. However, when operating CN or NFV services inevitably the question arises what an operator should do when the targeted efficiency and performance levels of virtualized infrastructures aren't met? How should an operator react when a customer complains about experienced performance and he doesn't want to sacrifice the efficiency of its infrastructure? Are hidden resources in a CN infrastructure available? Is a better load distribution possible? How is application-layer QoE related to the load seen on the host and in the network? How can a CN operator

identify efficiency leaks, performance bottlenecks or unused re-sources? Similar question was formulated within the REL (Reliability) working group of the ETSI ISG on NFV : how can an end-to-end and multi-level fault correlation be achieved? Due to the complexity of CN and NFV systems, current performance management methods are limited. Today's methods are mainly based on the current layering model. However, future methods neede to be capable to facilitate an end-to-end approach ("horizontal interoperability"; addressing all components along the service delivery chain, i.e. all server and networking elements) as well as a multi-layer concept for performance management ("vertical interoperability"; collaboration across all layers of the network stack and the virtualization layers).

We will present an initial approach and structuring on how to facilitate and to implement such an end-to-end multi-layer-based concept for performance management in CN and NFV infrastructure. In addition, we will discuss interoperability of load concepts as well as an initial correlation analysis, which enables the specification of relationships for load observed in guest VMs and the load observed on host (that offer resource sharing by virtualization techniques).

Kurt Tutschku is a professor in telecommunication systems at the Blekinge Institute of Technology (BTH) since March 2013. Before that, he was holding positions at University of Vienna (Chair of Future Communication; endowed by A1 Telekom Austria; Sep. 08 – Feb. 13), at NICT Japan (Feb. – Aug. 2008) and University of Würzburg, Germany (Head of Group on "New Network Architectures, Peer-to-Peer (P2P) Systems and Self-Organization"; 1999-2007). Kurt has received this Ph.D. and Habilitation degrees in CS in 1999 and 2008. His research interests include the architecture of the Future Internet (FI), network virtualization, Cloud Networking and the performance evaluation of future network control mechanisms. Kurt and his team have or are collaborating in various academic and industrial network virtualization projects such as FI-PPP (EU), GENI (US), or Akari (Japan). Kurt is also active in the ETSI ISG on Network Functions Virtualization (NFV).