State-of-the-Art and Challenges for COST Action

"Autonomous Control for a Reliable Internet of Services"

COST ACROSS, IC1304

Contents

1. Introduction

Today, we are witnessing a paradigm shift from the traditional information-oriented Internet into an Internet of Services (IoS). This transition opens up virtually unbounded possibilities for creating and deploying new services. Eventually, the Information and Communication Technologies (ICT) landscape will migrate into a global system where new services are essentially large-scale service chains, combining and integrating the functionality of (possibly huge) numbers of other services offered by third parties, including cloud services. At the same time, as our modern society is becoming more and more dependent on ICT, these developments raise the need for effective means to ensure quality and reliability of the services running in such a complex environment.

Motivated by this, the COST Action "Autonomous Control for a Reliable Internet of Services" (COST ACROSS) has been established to create a European network of experts, from both academia and industry, aiming at the development of autonomous control methods and algorithms for a reliable and quality-aware IoS.

About this document

The goal of this document is to identify the main scientific challenges to be faced during the course of the ACROSS project. To this end, we first give a general background and high-level description of the current state of knowledge. Then, for each of the three working groups we give a brief general introduction and background, and formulate a list of key research topics to be pursued during the Action's lifetime.

The document is the result of extensive interaction between the Working Group leaders, the Action Chairs and Management Committee.

This document gives an overview of the challenges that we are facing for the years to come. This document will be the basis for the research to be pursued in Y2 to Y4 for ACROSS.

2. General Background and Current State of Knowledge

The explosive growth of the Internet has fundamentally changed the global society. The emergence of concepts like service-oriented architecture (SOA), Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Cloud Computing has catalyzed the migration from the information-oriented Internet into an IoS. This has opened up virtually unbounded possibilities for the creation of new and innovative services that facilitate business processes and improve the quality of life. As a consequence, modern societies and economies have become heavily dependent on ICT. Failures and outages of ICT-based services (e.g., financial transactions, booking services, Web-shopping, governmental services) may cause economic damage and affect people's trust in ICT. Therefore, providing reliable and robust ICT services (resistant against system failures, cyber-attacks, high-load and overload situations, flash crowds, etc.) is crucial for our economy at large. Moreover, in the competitive markets of ICT service offerings, it is of great importance for service providers to be able to realize short time-to-market and to deliver services at sharp price-quality ratios. These observations make the societal and economic importance of reliable Internet services evident.

A fundamental characteristic of the IoS is that services combine and integrate functionalities of other services. This has led to complex service chains with hundreds - and often thousands - of services offered by different third parties, each with their own business incentives. In current practice, service quality for composite services is usually controlled on an *ad-hoc* basis, while the consequences of failures in service chains are not well understood. The problem is that, although such an approach might work for small service chains, this will become useless for future complex global-scale service chains.

Over the past few years, significant research has been devoted to controlling Quality of Service (QoS) and Quality of Experience (QoE) for IoS. To this end, much progress has been made in the functional layer of QoS-architectures and frameworks, and system development for the IoS. However, relatively little attention has been paid to the development, evaluation and optimization of algorithms for autonomous control. In this context, the main goal of this COST Action is to bring the state-of-the-art on autonomous control to the next level by developing quantitative methods and algorithms for autonomous control for a reliable IoS.

Traditionally, in the area of quantitative control methods the main focus has been on 'traditional' controls for QoS provisioning at the network and lower layers. In this context, it is important to note that control methods for the IoS also operate at the higher protocol layers and typically involve a multitude of administrative domains. As such, these control methods – and their effectiveness – are fundamentally different from the traditional control methods, posing fundamentally new challenges. For example, for composite service chains the main challenges are self-star methods for dynamic re-composition, to prevent or mitigate the propagation of failures through the service chains, and methods for overload control at the services layer.

Another challenging factor in quality provisioning in the IoS is its highly dynamic nature, imposing a high degree of uncertainty in many respects (e.g., in terms of number and diversity of the service offerings, the system load of services suddenly jumping to temporary overload, demand for cloud resources, etc.). This raises the urgent need for online control methods with self-learning capabilities that quickly adapt to – or even anticipate to – changing circumstances.

ACROSS will bring the state-of-the-art in the area of autonomous quality-based control in the IoS to the next level by ground-breaking research on developing efficient methods and algorithms that enable service providers to fully exploit the enormous possibilities for realizing a Research on quantitative methods for control in the IoS faces tremendous scientific challenges for which at best little is known today. ACROSS envisions the realistic situation where a huge number of composite service providers operate in an environment that is dynamic in different respects (e.g., availability of services, performance of services, user context, pricing schemes), where services may be composed of a large number of sub-services (including cloud services) offered by third parties. This requires research on the following important sub-areas:

1. Autonomous management and real-time control
2. Methods and tools for monitoring and service prediction
3. Smart pricing and competition in many-domain systems

These sub-areas are respectively covered by the three ACROSS Working Groups (WGs) WG1, WG2 and WG3. In the following sections we will elaborate on the scientific challenges faced in the context of each of these three WGs.

3. Challenges WG1 "Autonomous Management and Real-Time Control"

On a fundamental level, this WG is primarily concerned with the management and control of networks, services, applications, and compositions of services or applications. Of particular interest are management and control techniques that span multiple levels, e.g., the network and service level.

3.1 Introduction and background

To deliver reliable services in the IoS, service providers will need to implement control mechanisms, ranging from simplistic to highly advanced. Typical questions are: How can we realize the efficient use of control methods by properly setting parameter values and decision thresholds? How can we effectively use these mechanisms depending on the specific context of a user (e.g., in terms of user location, the user role, operational settings or experienced quality)? How do control methods implemented by multiple providers interact? How does the interaction between multiple control methods affect their effectiveness? What about stability? How to resolve conflicts? Ideally, control mechanisms would be fully distributed and based on (experienced) quality. However, some level of centralized coordination among different autonomous control mechanisms may be needed. In this context, a major challenge is to achieve a proper trade-off between fully distributed control (having higher flexibility and robustness/resilience) and more centralized control (leading to better performance under 'normal' conditions). This will lead to hybrid approaches, aiming to combine 'the best of two worlds'.

3.2 Research topics

In the following, we detail selected research challenges in the scope of WG1, which are being addressed by ACROSS members.

3.2.1 Control issues in softwarized networks

As part of the current cloud computing trend, the concept of cloud networking [Mell2011] has emerged. Cloud networking complements the cloud computing concept by enabling and executing network features and functions in a cloud computing environment. The supplement of computing capabilities to networks outlines elegantly the notion of "softwarization of networks". The added computing capabilities are typically general processing resources, e.g. off-the-shelf servers, which can be used for any computing requirement on any layer in the ISO/OSI stack model, i.e. on application layer (e.g. for the re-coding of videos) as well as network layer (e.g. for the computation of routes or the authentication of users). Hence, features and functions in the network-oriented layers are moved away from hardware implementations into software where appropriate, what is lately being termed as *network function virtualization (NFV)* [ETSI2014].

The Software-Defined Networking (SDN) paradigm [Ghod2011] emerged as a solution to the limitations of the monolithic architecture of conventional network devices. By decoupling the system that makes decisions about where traffic is sent (the control plane) from the underlying systems that forward traffic to the selected destination (the data plane), SDN allows network administrators to manage network services through abstraction of lower level and more fine-grained functionality. Hence, SDN and the softwarization of networks (NFV) stand for a "new and fine-grained split of network functions and their location of execution". Issues related to the distribution and coordination of software-based network functionality controlling the new simplified hardware (or virtualized) network devices forms a major research issue in ACROSS.

3.2.2 Scalable QoS-aware service composition using hybrid methods

Automated or semi-automated QoS-aware service composition is one of the most prevalent research areas in the services research community [Zeng2004, Fern2013, Leit2013a]. In QoS-aware composition, a service composition (or business process, or scientific workflow) is considered as an abstract graph of activities that need to be executed. Concrete services can be used to implement specific activities in the graph. Typically, there are assumed to be multiple functionally identical services with differing QoS available to implement each activity in the abstract composition. The instantiation problem is now to find the combination of services to use for each activity so that the overall QoS (for one or more QoS metrics) is optimal, e.g., find the services so that the QoS metric "response time" is minimal given a specific budget. The instantiation problem can be reduced to a minimization problem, and is known to be NP-complete.

Traditionally, QoS-aware service composition has been done using deterministic methods (e.g., simplex) for small service compositions, or a wide array of heuristics for large-scale problem instances (e.g., genetic algorithms, simulated annealing, and various custom implementations). However, the advent of cloud services and SDNs, service brokers, as well as the generally increasing size of service compositions requires new hybrid methods, which combine locally optimal solutions on various levels (e.g., on network, application, or service broker level). It is as of yet unclear how such optimizations on various levels, and conducted by various separate entities, can be optimally conducted and coordinated, and how stability of such systems can be ensured. However, one promising approach is the utilization of nature-inspired composition techniques, for instance the chemical programming metaphor.

3.2.3 Efficient use of multiple clouds

One of the challenges of current cloud computing systems is the efficient use of multiple cloud services or cloud providers. On network level, this includes the idea of virtual network infrastructures (VNIs), see e.g. [Jain2013]. The VNI concept assumes exploitation of network resources offered by different network providers and their composition into a common, coherent communication infrastructure supporting distributed cloud federation [Cont2013]. Controlling, managing and monitoring network resources would allow cloud federations to implement various new features that could: (1) optimize traffic between sites, services and users; (2) provide isolation for whole clouds or even for particular users e.g. that require deployment of their own protocols over the network layer; (3) simplify the process of extending and integrating CPs and NPs into the federation with reduced efforts and costs.

On service and application level, the idea of cloud bursting has been proposed as a way to efficiently use multiple cloud services [Guo2012, Leit2013b]. In cloud bursting, applications or services are typically running in a private cloud setup, until an external event (e.g., a significant load spike that cannot be covered with internal resources) forces the application to “burst” and move either the entire application or parts of it to a public cloud service. While this model has clear commercial advantages, its concrete realization is still difficult, as it cloud bursting requires intelligent control and management mechanisms for predicting load, deciding which applications or services to burst, and for technically implementing seamless migration. Additionally, the increased network latency is often a current practical problem in cloud bursting scenarios.

3.2.4 Energy-aware control

Traditionally, the optimization of ICT service provision made use of network performance (NP) related characteristics or key performance indicators as basic inputs for control and actuation loops. Those initially simple and technical only parameters evolved later to more complex QoE related aspects, leading to multivariate optimization problems. New control and actuation loops then involved several parameters to be handled in a joint manner due to the different tradeoffs and interdependencies among input and output indicators usually by composing the effects through a simplified utility function. Therefore, resulting approaches have put the focus particularly in the reward (in terms of users' satisfaction) to be achieved using efficiently the available network resources (see for instance, [Tabo2014]).

Meanwhile, the cost of doing so was most of the times faced as a constraint of the mathematical problem and considered again technical resources only. However, in the wake of “green ICT” and, more generally speaking, the requirement of economically sustainable and profitable service provision entail new research challenges where the cost of service provisioning must also consider energy consumption and price (for example, [Tabo2011]). The resulting energy- and price-aware control loops demand intensive research, as the underlying multiobjective optimization problem as well as the complexity of utility functions (see [Libe2013]) and the mechanisms for articulation of preferences exceed current common practices.

Such constraints affect not only network but the whole ICT service provision chain. For example, server farms are vital components in cloud computing and advanced multi-server queueing models that include features essential for characterizing scheduling performance as well as energy efficiency need to be developed. Recent results in this area include [Hyyt2014a,Hytt2014b,Gebr2014] and analyze fundamental structural properties of policies that optimize the performance-energy trade-off. On the other hand, several works exist [Rais2014,Dech2013] that employ energy-driven MDP solutions. In addition, the use of energy-aware multipath TCP in heterogeneous networks (see [Chen2013,Ding2014]) has become challenging.

3.2.5 New transport control protocols in the context of service compositions

There is a current trend to circumvent the tried-and-trusted TCP protocol by concepts like QUIC (a protocol from Google) and SPUD (an IETF initiative), see e.g. [Tram2014]. New DIY (do-it-yourself) protocols open a world of threats and opportunities. Threats include the risk of congestion collapse, as content providers develop more and more aggressive protocols and deploy faster and faster accesses in an attempt to improve their service, and inability to cache popular objects near users or prioritize between flows on congested access links as a result of the tendency to paint all traffic "grey", i.e. to encrypt even trivial things like public information. As for opportunities, TCP clearly has some performance problems and is a part of the ossification of the Internet. A (set of) new protocol(s) could circumvent the issues related to TCP and be adapted to present networks and content and therefore provide potentially better performance. Items of interest for research would thus include examining scenarios where caching or prioritization may be beneficial to identify cases where full-fledged encryption may be harmful, and to develop mechanisms like meta-information tags that allow for meaningful content caching and traffic discrimination and policing without infringing on, e.g., content copyrights or net neutrality, as well as researching approaches for congestion control. In the end, the goal will be to present new transport protocols that work better than present TCP, and at the same time compete with well behaved, legacy TCP in a fair way.

4. Challenges WG2 "Methods and Tools for Monitoring and Service Prediction"

4.1 Introduction and background

A crucial element for autonomous control in the IoS is monitoring and service prediction. For autonomous real-time (user-perceived) QoS and QoE in large, dynamic, complex many-domain environments like the IoS, there is a great need for scalable, non-intrusive monitoring and measurement of service demands, service performance and resource usage. Additional constraints regarding e.g. privacy and integrity will further complicate the challenges for monitoring and measurement. In addition, proactive service adaptation capabilities are rapidly becoming increasingly important for service-oriented systems like IoS. In this context, there is a need for online quality prediction methods in combination with self-adaptation capabilities (e.g., service re-composition). Service performance monitoring capabilities are also important for the assessment of Service Level Agreement (SLA) conformance, and moreover, to provide accurate billing information. In general, the metrics to monitor rely on the point of view adopted. For instance, cloud and service providers will need metrics to monitor SLA conformance and manage the cloud whereas composite service provider have to monitor multiple SLAs which is also different than what is required to be monitored for customers and service consumers.

4.2 Research topics

In the following, we detail selected research challenges in the scope of WG2, which are being addressed by ACROSS members.

4.2.1 How do define 'QoS' and 'QoE', what to measure?

A common definition of Quality of Experience (QoE) is provided in [LeCa2012]: "*QoE is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state.*" In contrast, ITU-T Rec. P.10 defines QoE as "*the overall acceptability of an application or service, as perceived subjectively by the end user*". The definition in [LeCa2012] advances the ITU-T definition by going beyond merely binary acceptability and by emphasizing the importance of both, pragmatic (utility) and hedonic (enjoyment) aspects of quality judgment formation. The difference to the definition of Quality of Service (QoS) by ITU-T Rec. E.800 is significant: *"[the] totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service*". Factors important for QoE like context of usage and user characteristics are not comprehensibly addressed by QoS.

As a common denominator, four different categories of QoE influence factors [LeCa2012,Hoss2013] are distinguished, which are influence factors on context, user, system, and content level. The context level considers aspects like the environment where the user is consuming the service, the social and cultural background, or the purpose of using the service like time killing or information retrieval. The user level includes psychological factors like expectations of the user, memory and recency effects, or the usage history of the application. The technical influence factors are abstracted on the system level. They cover influences of the transmission network, the devices and screens, but also of the implementation of the application itself like video buffering strategies. The content level addresses for the example of video delivery the video codec, format, resolution, but also duration, contents of the video, type of video and its motion patterns.

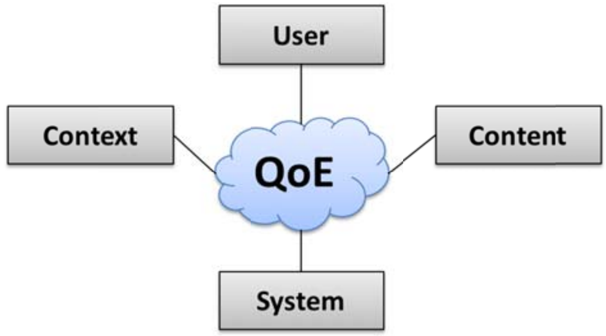

Figure 1: Different categories of QoE influence factors.

4.2.2 QoE and QoS monitoring for cloud services

The challenges of QoE management in cloud services are discussed in detail in [Hoss2012]. Cloud technologies are used to provision a whole spectrum of new and also traditional services. As users' experiences are typically application and service dependent, the generality of the services can be considered a big challenge in QoE monitoring of cloud services. Nevertheless, generic methods would be needed as tailoring of models for each and every application is not feasible in practice. Another challenge is brought up by multitude of service access methods. Nowadays, people use variety of different devices and applications to access the services from many kinds of contexts (e.g. different social situations and physical locations).

Traditional services that have been moved to clouds can continue using the proven existing QoE metrics. However, new QoS metrics related to the new kind of resources and their management (e.g. virtualization techniques, distributed processing and storage) and how they contribute to QoE require gaining new understanding. On the other hand, the new kind of services enabled by cloud (e.g. storage and collaboration) call for research regarding not only QoS to QoE mapping, but also the fundamentals on how users perceive these services. In addition to this, the much discussed security, privacy and costs need to be considered inside QoE topic.

4.2.3 QoE and context-aware monitoring

Today's consumer Internet traffic is transmitted on a best effort basis without taking into account any quality requirements. QoE management aims at satisfying the demands of applications and users in the network by efficiently utilizing existing resources. Therefore, QoE management requires an information exchange between application and network and proper monitoring approaches. QoE management requires three basic research steps, (1) modeling QoE, (2) monitoring QoE, and (3) optimizing QoE.

As a result of the QoE modeling process, QoE-relevant parameters are identified which have to be monitored accordingly. In general, monitoring includes the collection of information such as (1) the network environment (e.g., fixed or wireless); (2) the network conditions (e.g., available bandwidth, packet loss); (3) terminal capabilities (e.g., CPU power, display resolution); (4) service and application specific information (e.g., video bitrate, encoding, content genre). But also monitoring on application layer may be important, e.g. QoE monitoring for YouTube requires monitoring or estimating the video buffer status in order to recognize or predict when stalling occurs.

The QoE monitoring can either be performed a) at the end user or terminal level, b) within the network, or c) by a combination thereof. While the monitoring within the network can be done by the provider for fast reaction on degrading QoE, it requires mapping functions between network QoS and QoE. When taking into account application-specific parameters additional infrastructure like DPI is required to derive and estimate these parameters within the network. A better view on user perceived quality is achieved by monitoring at the end user level. However, additional challenges arise, e.g., how to feed QoE information back to the provider for adapting and controlling QoE. In addition, trust and integrity issues are critical as users may cheat to get better performance.

Going beyond QoE management, additional information may be exploited to optimize the services on a system level, e.g. resource allocation and utilization of system resources, resilience of services, but also the user perceived quality. While QoE management mainly targets at the optimization of current service delivery and currently running applications, the exploitation of context information by network operators may lead to more sophisticated traffic management, reduction of the traffic load on the inter-domain links, and a reduction of the operating costs for the ISPs.

Context monitoring aims at getting information about the current system situation from a holistic point of view. Such information is helpful for control decisions. For example, popularity of video requests may be monitored, events may be foreseen (like soccer matches) which allow to better control service and allocate resources. This information may stem from different sources like social networks (e.g. useful for popularity of videos and to decide about caching/bandwidth demands) but can also be monitored on the fly. Thus, context monitoring includes aspects beyond QoE monitoring. Context monitoring increases QoS and QoE (due to management of individual flows/users). But it may also improve the resilience of services (due to broad information about network "status").

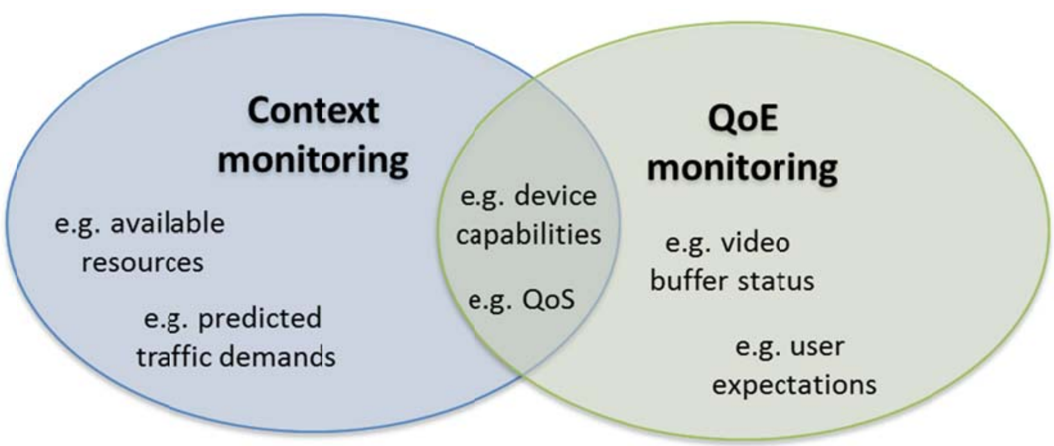

Figure 2: Relation between QoE monitoring and context monitoring

Context monitoring requires models, metrics and approaches which capture the conditions/system state of the system (network infrastructure, but also on service layer), but also application/ service demands, the capabilities of the end user device. The challenges are (1) the identification of relevant context information required for QoE but also reliable services, (2) quantification of QoE based on relevant QoS and context information, and (3) monitoring architecture and concept.

4.2.4 Predicting user needs for new services

Inevitably, Internet applications and services on a growing scale assist us in our daily life situations, fulfilling our needs for leisure, entertainment, communication or information. However, on the one hand, user acceptance of an existing Internet service/application depends on the variety of human factors influencing its perception, and, on the other hand, there are many human factors and needs, which could be supported by Internet Services and computing at large, yet are unknown to date. However, despite the importance of understanding of the human factors in computing, a sound methodology for evaluation of these factors and delineation of new ones, as well as reliable methods to design new Internet services with these factors in mind, do not exist.

This challenge goes beyond the QoE/QoS challenge presented in the previous section relating to a user experience with respect to an *existing and used* system. The challenge presented in here relates to identification of the unmet (implicit) needs of the user enabling *future* provision of novel and useful services. These human factors may relate to some specific phenomena ranging from, for example, the most preferred interaction style with a service (e.g., auditory, kinesthetic, visual) in a given context, via the user's specific health and care needs (e.g., wellness or anti-ageing), to the user's specific factors like cognitive load, physical flexibility, or momentary perception of safety, or intimacy in a specific context.

In this challenge we aim to provide a set of rigorous interdisciplinary, i.e., mixed-methods based methodological steps to be taken aiming to quantify human factors in computing in the user's natural environments and different contexts of service usage. The methodology incorporates qualitative and quantitative methods and involves real users in their real life environments through: (a) Gathering the cumulative users' opinion via *open-ended interviews and surveys*. Thus, specifically focusing on understanding the users' expectations towards a researched phenomena and current experience of this phenomena, mostly to establish the users' baseline experience on the experiment variables and context, but also to gather general demographics about the experiment participants. (b) Gathering the momentary users' opinion upon some specific factors like health behaviors, moods, feelings, social interactions, or environmental and contextual conditions via an *Experience Sampling Method* (ESM). Special momentary surveys executed multiple times per day 'in situ', i.e., in the natural users environments. (c) Gathering the episodic users' opinion upon some specific factors (as above) along semi-structured interviews based on the diary, for example by the *Day Reconstruction Method* (DRM).
(d) Gathering the data upon the users' daily life contexts and smartphone usage via continuous, automatic, unobtrusive data collection on the users' device through the measurements-based '*Logger*' service. Secondly, in this challenge we wish to provide guidelines on analyzing the relation of these factors with the design features of the computing system itself. Thirdly, we would like to provide

guidelines for Internet services/applications leveraging the human factors in their design process, assuring user's experience (QoE, linking to the previous section) and thus maximizing the user acceptance for these services.

4.2.5 End user monitoring; where to measure?

Quality monitoring is needed in all parts of the service delivery chain. However, sometimes the problems occur in the last mile, i.e. in access network, in end user's own networks or in user devices. In order to capture these problems, monitoring capabilities have to be included close to the user. Also regarding the composite product of the whole service machinery, the service that user sees, it is the end user monitoring that provides the most realistic view on perceived performance of service. In a single case, the implementation of such monitoring can be straightforward. However, the abundance of devices, platforms and applications used to access cloud services poses a challenge to implementing and deploying monitoring, which is further amplified by the lack of widely accepted standards for such monitoring. Getting the application developers, operators, service providers and all included parties to agree on end user monitoring standardization is one of the biggest challenges in this field. This topic should consider the QoE monitoring locations/points inside target systems and implications of different locations in terms of accuracy, scalability and (implementation) complexity.

4.2.6 Monitoring aggregated and encrypted data ('grey traffic')

Monitoring generally assumes that it is possible to extract from the data a set of parameters (e.g. fields within a packet) that enables to know what data (or services) is travelling (resp. provided). However there is recent tendency to paint all traffic "grey", i.e. to encrypt even trivial things like public information. Even though this may appear to protect user privacy, in fact such obfuscation complicates or prevents monitoring, caching and prioritization which could have been used to reduce costs and optimize user experience. Actually, it is not only the content that is being encrypted but also the protocol itself (i.e. only the UDP header or similar is open). This means that, contrary to present TCP, one cannot even monitor a flow in terms of data and acknowledgments to, e.g. detect malfunctioning flows (e.g., subject to extreme losses) or perform local retransmission (from a proxy, e.g.). Regarding content identification, the solution need not necessarily be unprotected content (there are reasons related to content ownership etc.) but one can imagine tags of different kinds. Then there is the challenge to find incentives that encourage correct labelling [Casa2014], such that traffic can be monitored and identified to the extent necessary to optimize networks (long term) and QoE (short term).

4.2.7 Timing- awareness for monitoring Future Internet infrastructures & applications

A key objective of ACROSS is to ensure that the ICT infrastructure that supports future Internet is designed such that the quality and reliability of the services running in such a complex environment can be guaranteed. This is a huge challenge with many facets, particularly as the Internet evolves in scale and complexity . One key building block that is required at the heart of this evolving infrastructure is precise and verifiable timing. Requirements such as 'RealTime Control', 'Quality Monitoring', 'QoS and QoE monitoring', 'Software Defined Networking' cannot easily or effectively be met without a common sense of precise time distributed across the full Infrastructure. ‘You cannot control what you do not understand’ is a phrase that applies here – and you cannot understand a dynamic and realtime system without having precise and verifiable timing data on its performance. As a first step – such timing services will firstly ensure that application and network performance can be precisely monitored, but secondly, and more importantly, will facilitate the design of better systems & infrastructure to meet the future needs. Unfortunately, current ICT systems do not readily support this paradigm. Applications, computers and communications systems have been developed with modules and layers that optimize data processing but often degrade accurate timing. State-of-the-art systems now use timing but usually only as a performance metric. To enable the massive growth predicted, accurate timing needs cross-disciplinary research to be integrated into existing and future systems. In addition timing accuracy, security of timing services issues represent another critical need. In many cases, having assurance that the time is correct is a more difficult problem than accuracy. A number of recent US initiatives are focusing on these challenges. See [TAAC2014] and [Cybe2014].

4.2.8 Prediction of performance, quality and reliability for composite services

Service composition structure contains important information about its performance, quality and reliability. It is well known that size, fault and failure distribution over software components in large scale complex software systems follow power law distributions [Hatt2009, Gali2014]. The knowledge of underlying generative models for these distributions enables developers to identify critical parts of such systems in early stages of development and act accordingly to produce higher quality and more reliable software at lower cost. Similar behavior is expected from large scale service compositions. The challenge is to extend the theory of distribution of various properties of services (e.g. performance) in a large scale service composition to be able to predict the behavior of such composition, determine critical services and service structures for the performance and reliability of the whole composition, evolution of service compositions and its quality attributes

4.2.9 Monitoring with Software Defined Networking

Software Defined Networking (SDN) is a new and promising networking paradigm [Mcke2008, Kreu2014]. It consists in decoupling control plane from forwarding plane and offers a whole set of opportunities to monitor the network performance. In SDN each node (router, switch, …) updates a controller about almost any information regarding the traffic traveling at any time in the network. A set of patterns can be defined by the controller for the node to apply this pattern and count the number of packets satisfying this specific pattern. Basic monitoring applies on the well-known header fields at any of 5 layers. However some Network Function Virtualization (NFV) can be introduced at some node to perform fine monitoring on the data (e.g. DPI to get specific info on data) and can therefore enable the controller to have full knowledge on what happens in the network.

Of course these great opportunities provided by SDN are accompanied by a list of (measurement and monitoring) challenges currently researched over the world [Haki2014,Jars2013,Chau2013]. For instance, how many controllers should we deploy? Too many controllers would bring us back to the older architecture but on the other hand too few controllers that centralize a large area would induce delay in getting the information and would require very expensive computation power from the controller to deal with huge amount of data. In this latter case this would also generate bottleneck near the area of the controller.

5. Challenges WG3 "Smart Pricing and Competition in Many-Domain Systems"

5.1 Introduction and background

Each service provider in the IoS will implement its own pricing mechanism, which may range from simple static to advanced dynamic pricing policies. The involvement of third-party and cloud services in making up a composite service in these dynamic and competitive environments (all of the involved parties striving for maximization of their own profit) raises challenging questions that are new, and moreover, one can learn from the past. For example, in the traditional Internet, flat-fee charging schemes tend to be replaced by volume-based charging schemes. In this context, typical questions are: What are the implications of implementing different pricing mechanisms in a many-domain setting? How do quality levels and pricing mechanisms relate? How can one develop smart pricing mechanisms that provide proper incentives for the involved parties (regarding brokering, SLA negotiation strategies, federation etc.) that lead to a stable ecosystem? What governing rules are needed to achieve this?

5.2 Research topics

In the following, we detail selected research challenges in the scope of WG3, which are being addressed by ACROSS members.

5.2.1 Modeling QoS/QoE-aware pricing issues

A key challenge is to understand what are the correct digital "goods" (e.g. in the cloud, in a distributed setting, beyond just physical resources), and at what level of granularity to consider pricing and competition issues [Abhi2012, Bhat2011]. Hence a framework (language /ontology) is needed for specifying and describing

- the digital goods and services for an IoS;
- a QoS and QoE framework for describing services;
- component specification that allows services to be built up from components;
- brokering, transfer charging and "exchanges" to allow for third parties, and for multi-provider services.

5.2.2 Context-dependent pricing, charging and billing of composite services

Pricing, charging and billing of composite services, provided to the end user by different service providers (SPs), requires the construction and elaboration of new mechanisms and techniques in order to provide the best service to the user [Ji2011,Vale2013,Abun2014], depending on their current context[1], and to enable viable business models for the service providers [Odro2010,Bloc2014,Gao2012]. Solving this problem requires advances in mechanism design: current economic theory lacks the sophistication to handle the potentially rich variety of service descriptions and specifications that could be delivered in the IoT/Next Gen Internet.

Charging and billing (C&B) requires mechanisms to allow secure transactions, trusted third party (TTP) C&B [Jaka2011], cross-provider payments, micropayments and to allow for new payment paradigms, such as peer-to-peer currencies. The TTP feature of the C&B entity, perhaps, will also facilitate the initial establishment of trust and subsequent interaction, e.g. to ensure interoperability, between different SPs as regards the services (service components) provided by each of them.

The pricing and C&B need to be aligned with service definition and implementation. Hence the Autonomous Controls of WG1 need to be inextricably linked to pricing, and what can be measured (WG2). This challenge relates also to the service IDS (intelligent demand shaping) and service MAP (measurement, analytics, profiling). As a specific example, service delivery and SLAs are linked to the dynamic monitoring of the quality of each component of the composite service, with an ability to dynamically replace/substitute the component(s) that is/are currently underperforming with another one(s), which is/are identified as working better in the current context. The replacement of service components must be performed transparently to the user – perhaps with the user only noticing improvements in the overall service quality.

5.2.3 QoS and price-aware selection of cloud service providers

The uprise of Infrastructure-as-a-Service (IaaS) clouds has led to an interesting dilemma for software engineers looking to adapt the cloud. Fundamentally, the basic service offered by different providers (e.g., Amazon EC2, or, more recently, Google Compute Engine) is entirely interchangeable. However, non-functional aspects (e.g., pricing models, expected performance of acquired resources, stability and predictability of performance) vary considerably, not only between providers, but even among different data centers of the same provider. This is made worse by the fact that, currently, IaaS providers are notoriously vague when specifying details of their service (e.g, "has two virtual CPUs and medium networking performance"). As a consequence, cost-aware cloud users are currently not able to make an informed decision about which cloud to adopt, and which concrete configuration (e.g, instance type) to use for which application. Hence, cloud users often base their most fundamental operations decisions on hearsay, marketing slogans, and anecdotal evidence rather than sound data. Multiple research teams worldwide have proposed tools to allow developers to benchmark cloud services in a more rigid way prior to deployment (e.g, CloudCrawler, CloudBench, or Cloud Workbench, [Sche2014,Frey2013]).

[1] The *context*, from which the price is computed, has three main components – *user context* (i.e. the user location, the user preferences and profile(s), the user mobile device(s), etc.), *network context* (i.e. the network congestion level, the current data usage pattern of the network, the current QoS/QoE index of the network, the cost of using the network, etc.), and *service context* (i.e. the category, type, the scope and attributes of the service being requested, the request time, the application initiating the request, the current QoS/QoE index of the service component, etc.).

However, so far, we are missing fundamental insight into which kind of IaaS provider and configuration is suitable for which kind of application and workload.

6. References

[Abhi2012] Vineet Abhiskeh, Ian A. Kash and Peter Key, Fixed and Market Pricing for Cloud Services. In: Proceedings 7th Workshop on Economics of Networks, Systems and Computation (NetEcon 2012), Orlando, USA, March 2012.

[Abun2014] M. Abundo, V. Di Valerio, V. Cardellini, F. Lo Presti, Bidding strategies in QoS-aware cloud systems based on N-armed bandit problems. In: Proceedings of the 3rd IEEE Symposium on Network Cloud Computing and Applications (NCCA 2014), pp. 38-45, Feb. 2014.

[Bhat2011] S. Bhattacharjee, R. D. Gopal, J. R. Marsden and R. Sankaranarayanan, Digital Goods and Markets: Emerging Issues and Challenges. ACM Transactions on Management Information Systems, Vol. 2. No. 2, 2011.

[Bloc2014] Gideon Blocq, Yoram Bachrach, and Peter Key, The Shared Assignment Game and Applications to Pricing in Cloud Computing. In Proceedings AAMAS 2014, Paris, France, May 2014.

[Casa2014] John Casaretto, "The Internet strikes back: Global encrypted SSL traffic booms". http://siliconangle.com/blog/2014/05/20/the-internet-strikes-back-global-encrypted-ssl-traffic-booms/, 2014.

[Chau2013] C. Chaudet, Y. Haddad, Wireless Software Defined Networks: Challenges and opportunities. IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems (COMCAS 2013), pp.1-5, Oct. 2013.

[Chen2013] S. Chen, Z. Yuan and G. M. Muntean, A traffic burstiness-based offload scheme for energy efficiency deliveries in heterogeneous wireless networks, In IEEE Globecom Workshops, Atlanta, USA, 2013.

[Cont2013] CONTRAIL project (Seamless Cloud federation through open virtualization infrastructure), EU 7th Framework Programme, http://contrail-project.eu/, 2010-2013.

[Cybe2014] Cyber-Physical Systems Public Working Group (CPS PWG), http://www.cpspwg.org/, 2014.

[Dech2013] D. Dechen and A. Shami, Energy efficient QoS constrained scheduler for SC-FDMA uplink, Personal Communication, vol. 8, pp. 81-90, 2013.

[Ding2014] T. Ding et al., Smartphone energy consumption models for multimedia services using multipath TCP, 11th IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, USA, 2014.

[ETSI2014] ETSI ISG on NFV: "Network Functions Virtualisation – White Paper #3", Available at http://portal.etsi.org/NFV/NFV_White_Paper3.pdf, Oct 2014.

[Fern2013] Héctor Fernández and Cedric Tedeschi, A Chemistry-Inspired Workflow Management System for Decentralizing Workflow Execution. IEEE Int. Journal Transactions on Service Computing, vol.99, 2013.

[Frey2013] Sören Frey, Florian Fittkau, and Wilhelm Hasselbring, Search-based genetic optimization for deployment and reconfiguration of software in the cloud. In: Proceedings of the 2013 International Conference on Software Engineering (ICSE '13). IEEE Press, Piscataway, NJ, USA, pp. 512-521, 2013.

[Gao2012] Xi Alice Gao, Yoram Bachrach, Peter Key, and Thore Graepel, Quality Expectation-Variance Tradeoffs in Crowdsourcing Contests. In: Proceedings AAAI 2012, Toronto, Canada, July 2012.

[Gebr2014] M. E. Gebrehiwot, S. Aalto and P. Lassila, Optimal sleep-state control of energy-aware M/G/1 queues, in Proceedings of 8th International Conference on Performance Evaluation Methodologies and Tools (Valuetools), Bratislava, Slovakia, 2014, to appear.

[Gali2014] Tihana Galinac Grbac, Darko Huljenić, On the Probability Distribution of Faults in Complex Software Systems. Information and Software Technology, published online first, 2014.

[Ghod2011] Ali Ghodsi, P. Brighten Godfrey, Nick McKeown, Guru Parulkar, Barath Raghavan, Architecting for Innovation. ACM Computer Communications Review, 2011.

[Guo2012] Tian Guo, Upendra Sharma, Timothy Wood, Sambit Sahu, and Prashant Shenoy, Seagull: intelligent cloud bursting for enterprise applications. In: Proceedings of the 2012 USENIX Annual Technical Conference (USENIX ATC'12), USENIX Association, Berkeley, CA, USA, 2012.

[Haki2014] A. Hakiri et al., Software-Defined Networking: Challenges and research opportunities for Future Internet. Computer Networks (in press), available online http://www.sciencedirect.com/science/article/pii/S1389128614003703#, October 2014.

[Hatt 2009] Les Hatton, Power-Law Distributions of Component Size in General Software Systems. IEEE Trans. Software Eng. 35(4), pp. 566-572, 2009.

[Hoss2013] Hossfeld, Tobias, Raimund Schatz, Ernst Biersack, and Louis Plissonneau. "Internet video delivery in YouTube: from traffic measurements to quality of experience." In *Data Traffic Monitoring and Analysis*, pp. 264-301. Springer Berlin Heidelberg, 2013.

[Hoss2012] T. Hossfeld, R. Schatz, M. Varela, and C. Timmerer, Challenges of QoE management for cloud applications, IEEE Communications Magazine, vol. 50, no. 4, pp. 28–36, 2012.

[Hyyt2014a] E. Hyytiä, R. Righter and S. Aalto, Task Assignment in a Server Farm with Switching Delays and General Energy-Aware Cost Structure, Performance Evaluation, vol. 75-76, pp. 17-35, 2014.

[Hytt2014b] E. Hyytiä, R. Righter and S. Aalto, Energy-aware Job Assignment in Server Farms with Setup Delays under LCFS and PS, in 26th International Teletraffic Congress (ITC'26), Karlskrona, Sweden,2014.

[Jain2013] [JaKM13] Sushant Jain, Alok Kumar, Subhasree Mandal et al., B4: experience with a globally-deployed software defined WAN. In: Proceedings of SIGCOMM 2013, ACM, New York, USA, 2013.

[Jaka2011] J. Jakab, I. Ganchev, M. O'Droma, "Third-Party Charging and Billing for the Ubiquitous Consumer Wireless World". International Journal on Communications Antenna and Propagation, Vol. 1, No. 2, pp. 136-144, 2011.

[Jars2013] M. Jarschel et al., On the Accuracy of Leveraging SDN for Passive Network Measurements. In: Proceedings ATNAC 2013, Christchurch, New Zealand, November 2013.

[Ji2011] Z. Ji, I. Ganchev, M. O'Droma, "An iWBC Consumer Application for 'Always Best Connected and Best Served': Design and Implementation". IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, pp. 462-470, 2011.

[Kreu2014] D. Kreutz et al., Software-defined networking: A comprehensive survey. Online available: http://arxiv.org/abs/1406.0440, 2014.

[LeCa2012] Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.

[Leit2013a] Philipp Leitner, Waldemar Hummer, and Schahram Dustdar, Cost-Based Optimization of Service Compositions. IEEE Trans. Serv. Comput. 6, 2, pp. 239-251, 2013. http://dx.doi.org/10.1109/TSC.2011.53

[Leit2013b] Philipp Leitner, Zabolotnyi Rostyslav, Alessio Gambi, and Schahram Dustdar, A Framework and Middleware for Application-Level Cloud Bursting on Top of Infrastructure-as-a-Service Clouds. In: Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing(UCC'13). IEEE Computer Society, Washington, DC, USA, 163-170, 2013. http://dx.doi.org/10.1109/UCC.2013.39

[Libe2013] F. Liberal, I. Taboada and J. O. Fajardo, Dealing with Energy-QoE Trade-Offs in Mobile Video, Journal of Computer Networks and Communications, vol. 2013.

[Mcke2008] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner, OpenFlow: enabling innovation in campus networks. SIGCOMM Comput. Commun. Rev. 38, 2, pp. 69-74, 2008.

[Mell2011] Peter Mell, Timothy Grance, The NIST Definition of Cloud Computing. National Institute of Standards and Technology, NIST Special Publication 800-145, Sept. 2011.

[Odro2010] M. O'Droma and I. Ganchev, "The Creation of a Ubiquitous Consumer Wireless World through Strategic ITU-T Standardization". IEEE Communications Magazine, Vol. 48, Issue 10, pp. 158-165, 2010.

[Rais2014] M. Raiss-el-fenni et al, POMDP game framework for service providers inciting mobile users, IEEE International Conference on Communication (ICC), Sydney, Australia, 2014.

[Sche2014] Joel Scheuner, Philipp Leitner, Juergen Cito, and Harald Gall, Cloud WorkBench – Infrastructure-as-Code Based Cloud Benchmarking. In: Proceedings of the 6th IEEE International Conference on Cloud Computing Technology and Science (CloudCom'14), 2014.

[TAAC2014] TAACCS: Cross-Disciplinary Research on Time-Aware Applications, Computers, and Communication Systems, http://www.taaccs.org/index.html, 2014.

[Tabo2011] I. Taboada, J. O. Fajardo and F. Liberal, QoE and energy-awareness for multi-layer video broadcasting, In IEEE International Conference on Multimedia and Expo (ICME), Barcelona , Spain, 2011.

[Tabo2014] I. Taboada and F. Liberal, A Novel Scheduling Index Rule for QoE Maximization in Wireless Networks, Abstract and Applied Analysis, vol. 2014.

[Tram2014] Brian Trammell and Joe Hildebrand, "The Trouble with Transports - Moving UDP to layer 3.5". http://www.ietf.org/proceedings/90/slides/slides-90-appsawg-7.pdf, 2014.

[Vale2013] V. Di Valerio, V. Cardellini, F. Lo Presti, Optimal pricing and service provisioning strategies in cloud systems: a Stackelberg game approach. In: Proceedings of the IEEE 6th International Conference on Cloud Computing (CLOUD 2013), Santa Clara, CA, pp. 115-122, June 2013.

[Zeng2004] Liangzhao Zeng, Boualem Benatallah, Anne H.H. Ngu, Marlon Dumas, Jayant Kalagnanam, and Henry Chang, QoS-Aware Middleware for Web Services Composition. IEEE Trans. Softw. Eng. 30, 5, pp. 311-327, 2004. http://dx.doi.org/10.1109/TSE.2004.11