The use of NGN / IMS for Cloud and Grid Services Control and Management

Fabricio Carvalho de Gouveia[2], Richard Good[1], Thomas Magedanz[2], Neco Ventura[1]
[1]University of Cape Town, Rondebosch, 7700, South Africa
{rgood | neco}@crg.ee.uct.ac.za
[2]Technische Universitaet Berlin / Fraunhofer Institute FOKUS
Franklinstr. 28-29, D-10587
fabricio.gouveia@fokus.fraunhofer.de | tm@cs.tu-berlin.de
Faculty IV / Next Generation Networks
Berlin, Germany

Abstract—The ETSI Technical Committee (TC) GRID is addressing issues associated with the convergence of IT and Telecommunications, with particular reference to the lack of interoperable Grid solutions. The integration of grid technology provides new revenue streams for Telecom operators through their converged fixed, mobile and data services from their Next Generation Network (NGN) and by offering "Grid and cloud computing data" services reusing the current deployed NGN infrastructure. TC GRID identified the gaps and overlaps that need to be addressed for this interconnection. This paper analyses two interconnection scenarios for combining grid and NGN applications in a unified architecture. The proposal takes advantage of standardized interfaces and current technology features for integrating NGN and Cloud Services. The use of the TISPAN Resource and Admission Control System (RACS) and Policy and Enforcement Manager (PEEM) are described as possible vehicles for the integration.

Index Terms—Cloud Computing, IP Multimedia Subsystem, Next Generation Networks, Unified.

I. INTRODUCTION

OVER the Top (OTT) architectures are proprietary solutions where an Application Server (AS) offers a service to clients through an IP pipe. This approach offers fast one-off deployment and optimized one-off capital expenditure (CAPEX). The problem lies in the individual operational expenditure (OPEX) required for each stand alone solution, which can lead to a fragmented end-user experience. For the deployment of new services, separate application servers need to be deployed and their respective CAPEX and OPEX repeated.

Next Generation Networks (NGN) decouple services from the transport plane in order to provide flexibility with a common control architecture. An example framework is the IP Multimedia Subsystem (IMS), where a core network is introduced to manage multimedia resources with different Quality of Service (QoS), security and charging models. This NGN approach offers faster integration of subsequent applications with the advantage of reusing the current deployed infrastructure. OPEX costs are shared across the whole architecture. End-users have the benefit of enhanced experience through integrated services.

Telecoms & Internet converged Services & Protocols for Advanced Networks (TISPAN) specify standards for fixed networks and Internet convergence, also known as Fixed Mobile Convergence (FMC), providing the definition of the NGN [1]. The NGN specifications are referenced by the ITU-T and TISPAN is considered the global NGN reference architecture. In its Next Generation Network Architecture (Fig. 1) a generic multi-service, multi-protocol, multi-access IP-based framework is presented that aims to become the reference model to achieve convergence between PSTN and IP data networks. This model is based on the concept of cooperating subsystems (SS) sharing common components. This architecture enables the smooth addition of new subsystems to cover demands and service classes, and ensures maximum common usage of network resources, applications and user equipment.

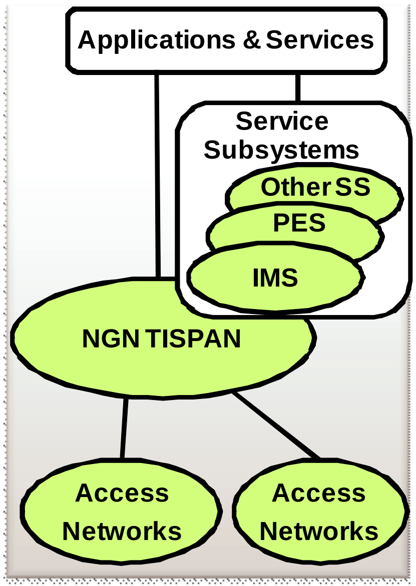

Fig. 1. NGN Reference Architecture.

A critical subsystem in this context is the IMS, which is a technology designed to provide robust multimedia services across roaming boundaries and diverse access technologies with guaranteed QoS, charging models, reliability and security.

NGN and IMS offer unified management and control, an all IP environment for improved performance and reduced network complexity. Flexibility, necessary for new services, is provided through common service capabilities all through standardized interfaces and Applications Programming Interfaces (APIs).

In the NGN the Resource and Admission Control System (RACS) plays an important role for access control. It is an independent subsystem in the TISPAN architecture, and can support QoS requests from other service subsystems (including IP Multimedia Subsystem), prevents congestion and supports transport layer mechanisms.

In the application layer, the Policy Evaluation, Enforcement and Management (PEEM) is responsible for authorizing and controlling the behavior of service requests. This is done through policies that define what resources can be accessed and by whom.

This paper discusses the integration of NGN/IMS and Cloud environments (Fig. 2). In particular the advantages of such an integration are outlined and architectural options are presented. These architectures have specific design features that when combined will provide mutual benefits and define a new paradigm for Cloud service management.

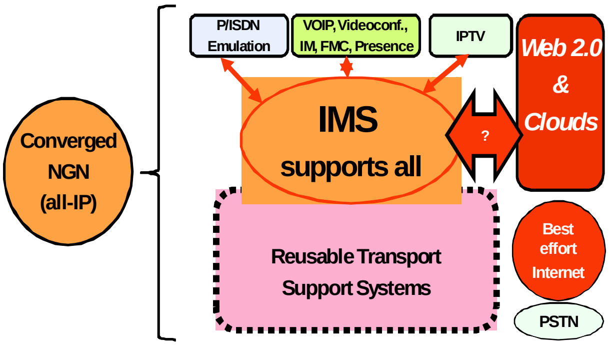

Fig. 2. Clouds as new NGN Application Domain?

The remainder of the paper is organized as follows. Section II describes the concepts of grid applications and cloud computing. Section III identifies gaps in the TISPAN standardisation for the integration of NGN and Cloud environments. In Section IV architectural scenarios for a unified architecture are discussed. Section V presents an impact analysis for a unified NGN and Cloud architecture. Section VI concludes the paper and discusses future work directions.

II. CLOUD/GRID BACKGROUND

The definitions of cloud and grid computing are closely related, open to interpretation and often misused. In this section we provide a definition for these technologies and describe how they relate to one another.

A. *Grid Computing*

Grid computing is a concept based on the collaboration of people and resources. This concept uses several computers to enable the sharing, selection, and aggregation of a wide variety of geographically distributed computational resources (e.g. computer clusters, storage systems). The system behaves like a single, unified resource for supporting large-scale and data-intensive computing applications [4].

It works as a parallel processing architecture with CPU resources shared across the network, to functions as a single supercomputer.

Grid computing has several promising features:

- It can make more cost-effective use of a given amount of computer resources.
- It can solve problems that require huge amounts of computing power without the need for specialized, expensive hardware.
- It allows the resources of many computers to be cooperatively and synergistically harnessed and managed as collaboration toward a common objective.

B. *Cloud Computing*

Cloud computing refers to a paradigm shift where computing is moved away from personal computers or an individual application server to a cloud of computers [10]. End-users of a cloud service need not be aware of the underlying details of how the computing and processing occurs.

These underlying details are what tie the cloud and grid computing environments together. In cloud computing, computer resources are pooled together and managed by software that can be based on the distributed grid computing model. Essentially cloud computing has evolved from grid computing. Grid computing typically refers to a single high performance application that is possibly batch scheduled, while cloud computing is often transactional and offers a variety of on-demand services [3].

The cloud computing concept incorporates the delivery of Infrastructure, Platform and Software as a Service [11]. Besides these categories, cloud services are offered directly to the users. These can be Internet-based services, such as middleware, database and storage. Cloud computing intends to offer resources with the following characteristics:

- Flexibility;
- Abstracted resources featuring scalability;
- Pay as you use model;
- Reliability;
- Performance.

The cloud architecture comprises hardware and software that communicate with each other over application programming interfaces, usually web services. Complexity is controlled and the resulting systems are more manageable than their monolithic counterparts.

The cloud architecture extends to the client, where web browsers and/or software applications access cloud applications. The two major forms of cloud services are compute clouds and storage clouds. The Amazon Elastic Computing Cloud (EC2) and Amazon Simple Storage Service (S3) are the pioneering applications in this sphere [8], [9]. It has been concluded that convergence of competing IT paradigms is needed in order to deliver the 21st century vision [12]. Throughout the remainder of this paper we refer to a Grid/Cloud domain that provides cloud services and manages shared computer resources using the grid computing model.

III. PROBLEM STATEMENT

In the current telecommunications environment of dwindling traditional revenues where OTT services are relegating operators to bit pipe suppliers, cloud computing provides a lucrative opportunity. These services promise to at least increase network utilization and thus transport revenues. Furthermore operators can realistically extract two revenue streams from the same function, charging both end-users and cloud service operators for a given service quality [6].

However it is the NGN features that provide the most significant opportunities for cloud computing. The All-IP infrastructure and highly modular architecture of the NGN would allow each cloud service to be treated as a hosted application. Furthermore the existing IMS enablers including capability negotiation, authentication, service invocation, addressing, routing, group management, presence, provisioning, session establishment, and charging, could be provided via standardized interfaces to the Grid/Cloud domain. Cloud service providers could also benefit from the unified management and control architecture.

It is clear that a unified IMS/NGN and Grid/Cloud domain has advantages for all stakeholders including end-users, network operators and cloud service providers. However there are architectural challenges for such a unified architecture. The ETSI Technical Report 102 659-2 considers barriers to interoperability between grid technologies and the NGN architecture and identifies standardization gaps that must be addressed to ensure successful roll out of said technologies [3].

Five areas are covered in detail: Architecture, Service Level Agreements, Charging, Security, and Service Discovery. In the presence of a multiplicity of network technologies and the resulting integration problems, the vertical integration of network layers is increasingly important. These architectural elements are critical to the NGN and to the formation and operation of dynamic large-scale grid infrastructures.

Current gaps highlighted in the report include:

- Support of end-to-end service on Grid/Cloud, NGN and in a mixed Grid/Cloud and NGN environment is not available.
- Mapping of different reference models between the NGN and Grid/Cloud environments is elementary.

The architecture must support end-to-end services, where the quality of network services requested by the grid layer needs to be independent of the underlying networking technologies. It is the responsibility of the network service provider to map and enforce the required quality of service. Currently neither NGN nor Grid/Cloud domains provide suitable interfaces or models to manage this relationship.

Horizontal integration of grid and NGN architectures needs to support the co-existence of multiple network service providers for widely distributed grid applications. These providers need to allow collaborative mechanisms for end-to-end service establishment. Current cross-network standards focus primarily on network provider interfaces and relatively static topologies. To realize the full potential of an integrated NGN and Grid/Cloud environment it is necessary to expose the cross-network routing and QoS interfaces to third-party applications for real-time dynamic service provisioning. A further major shortcoming of the Grid/Cloud standards landscape is the lack of a widely agreed upon architectural reference model.

To address these issues it is necessary to abstract the Grid/Cloud domain and view it as a diverse resource that can be controlled by core NGN components over standardized interfaces. Grid resources like computing power, network and storage resources can be managed in a similar fashion to the NGN transport plane via the Resource and Admission Control Subsystem (RACS). The question remains how to integrate these standardized interfaces into the Grid/Cloud environment. Furthermore the IMS services and enablers should be exposed to the Cloud environment allowing them to be reused and cloud services should be made available to the IMS application domain.

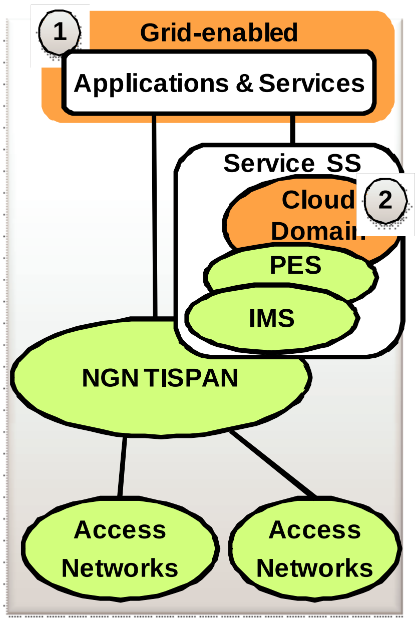

Fig. 3. Unified NGN Cloud/Grid Reference Architecture.

IV. UNIFIED ARCHITECTURE

There are two primary architectural scenarios for combining NGN and Grid technologies [3]. These are: Applications implemented with Grid technology and a Grid/Cloud architecture implemented as an NGN subsystem. Fig. 3 demonstrates these different approaches to forming a unified architecture.

A. Scenario 1: Grid-enabled applications and services for NGN

In terms of network requirements Grid applications are diverse. While it is not possible to define a generic Grid application, it is likely that they will be characterized by intensive use of IT resources, whether this is computational, access to large datasets or demanding constraints on data volumes, request rates or latency, transactional throughput or some combination of these.

However they are all characterized by a need for network connectivity and hence can be considered NGN applications. Grid enabled applications run within a Grid

Application Server, which has access via standardized interfaces to existing subsystems, e.g. ISC interface between IMS and AS. The impact that this coordination has on the defined NGN reference points and the architecture as a whole is under investigation by standardization bodies.

B. Scenario 2: Cloud Architecture as NGN Subsystem

The second approach implements a dedicated service control architecture to support Grid services. This Grid service subsystem provides access to grid resources, through resource virtualization and offers a service interface. This facilitates three use cases:

1. Grid services can be offered to Grid-enabled applications.
2. Grid resources can be offered to end-user applications.
3. Dedicated grid-enabled functions from other subsystems can interact with the Grid subsystem.

These scenarios are defined at a high level and detailed architectural information is not provided. In the following section they are expanded upon to define a unified architecture that combines grid and networking resources. This architecture uses existing standardized reference points and functional elements to facilitate the interaction. IMS and Grid services are made available to IMS and Cloud domains; Grid resources can be reserved in a standardized manner via the RACS and similarly grid applications can reserve resources in the transport plane.

V. IMPACT ANALYSIS FOR NGN-CLOUD INTEGRATION

Provisioning of Grid/Cloud services with the functional features of NGN requires a cooperated framework for interworking all players involved in the delivery process. This new framework will be subject to additional requirements and challenges. The players have to interact smoothly to fulfill the task of providing a new service experience to the end user. Building upon the capabilities of IMS as an overlay control subsystem designed to support heterogeneous IP networks and its ability to deliver integrated voice and data services while providing standardized interfaces, it will be used as the foundation convergence platform for the two scenarios identified in section 4. The first proposal involves the use of the OMA defined Policy Evaluation, Enforcement and Management (PEEM) framework to control access to and compose services on demand.

A. PEEM enabler for controlling Grid / Cloud services

New services developed by 3rd Party service providers (e.g., Web Services, Grid Services) create opportunities for the mobile industry. New value-added services can be created by combining diverse service enablers. These services are typically personalized, hence user and business partner information is needed during the service creation process. User’s information, business partner information and network resources have to be protected. One needs a central point that enforces service composition, orchestration, delegation and also controls access to enablers, applications and network resource.

The Open Mobile Alliance (OMA) defined PEEM was designed in order to provide a mechanism for exposure of 3rd party providers’ services to different clients in a controllable and personalized manner [6].

The PEEM system builds a gateway that provides controlled access, service personalization and resource reutilization, facilitating cooperation between 3rd party application providers and mobile industry operators that are interested in providing highly innovative services.

Fig. 4 depicts the context of PEEM for coping with the Scenario 1 requirements described in Section 3.

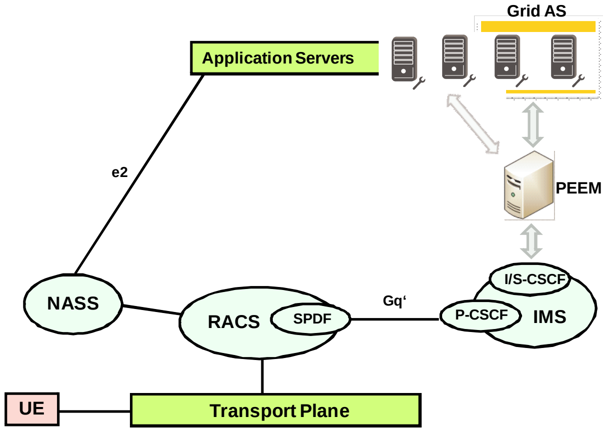

Fig. 4. The use of PEEM to offer Third Party Service Control to the NGN Domain.

Trusted applications and/or 3rd party applications have access to network resources only by triggering actions in the communication networks through the PEEM enabler. In order to achieve this, PEEM evaluates policies associated with each incoming request and enforces the decision result. This decision consists of authorization of the issued request and, in some cases depending on the policy, of requests to shared services/applications’ capabilities (e.g. charging, logging capabilities).

These features motivate the use of the PEEM to control service requests from Grids to the NGN or vice versa. With specifically designed policies PEEM can take the correct action when controlling and authorizing the Grid and Cloud services with included IMS enablers like location and presence. Figure 5 depicts the role players with which PEEM interacts: service requests (e.g. from applications) that initiate requests for capabilities exposed by the service provider, services and enablers intermediated by PEEM.

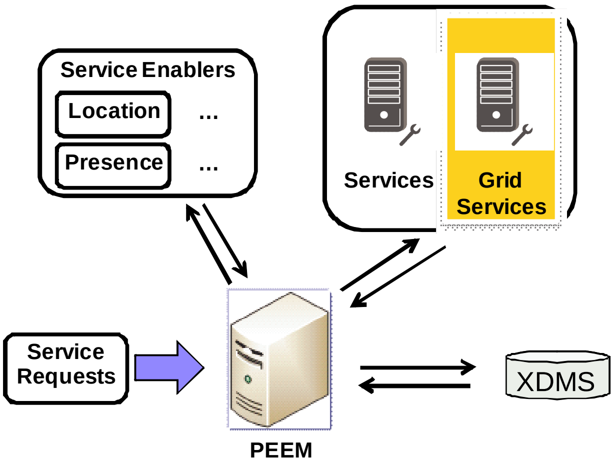

Fig. 5. PEEM role players.

PEEM exposes two kinds of interfaces, one for requests/responses for policy evaluation and another for management. PEEM can facilitate dynamic management and controlled access to Cloud resources through the usage of policies and the orchestration of Cloud resources (services and infrastructure resources) through the usage of common exposed APIs. This information can be exposed to service registries, and APIs that support the description of resource dependencies and interconnections for dynamic and reliable composition.

With the PEEM approach, large-scale data processing (a key characteristic of cloud architectures) can be offered by Telecommunication providers to increase revenues and increasing user usage options.

B. *RACS for controlling a Grid / Cloud Subsystem*

The usage of resources like computing power and storage in grid services is analogous to transport plane resources, like bandwidth, that are consumed by IMS sessions. Requests from grid applications can be thought of as similar to authorization and resource requests initiated by IMS application servers. The management of these resources should be flexible and under the control of the operator, preferably managed by static and dynamic policies.

The Resource and Admission Control Subsystem (RACS) is defined as a functional element within the NGN to expose resource management functions to applications [7]. The RACS utilizes policy based management to control access to resources. Applications compose authorization requests; the RACS authorizes requests based on policies that can be static or dynamic, and reserves these resources by configuring the physical devices in the transport plane. This architecture defines in-depth control scenarios and protocol specifications.

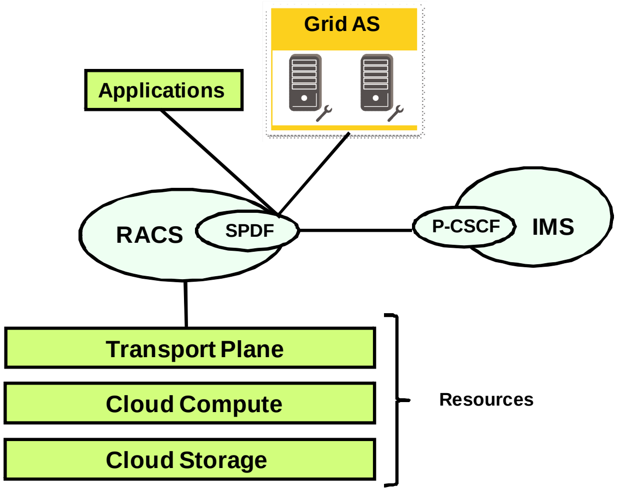

Fig. 6. RACS exposes resources to grid and IMS applications.

It is a compelling notion to reuse these standardized principles to expose the resources inherent in grid applications (computing power, storage, etc.) to IMS applications and manage this access through pre-defined policies. This would take advantage of already existing infrastructure and facilitate inter-operability between the NGN and Grid/Cloud architectures. The resulting unified architecture could manage all available resources towards specific applications in a flexible and generic manner.

Fig. 6 illustrates the extended scope of the RACS framework. The existing interface to the application plane facilitates resource requests from all service domains including grid applications. Based on the standardized policies the RACS performs authorization, reservation and commitment of resources. Policy rules are created that reserve the resources. These rules are generic and depending on the RACS policies, they result in the reservation of transport plane, cloud storage or cloud compute resources. It is important to note that this architecture exposes transport plane resources to grid applications, and grid resources to grid and IMS applications. This exposure is managed by the standardized mechanisms and policies of the RACS.

Fig. 7 shows the role players in the extended interactions. The RACS receives resource requests from applications (Grid, IMS, etc.). Policies are retrieved from the XDMS and authorization takes place. Generic policy rules are created and result in resources being assigned. These resources can be transport plane, cloud compute or cloud storage resources.

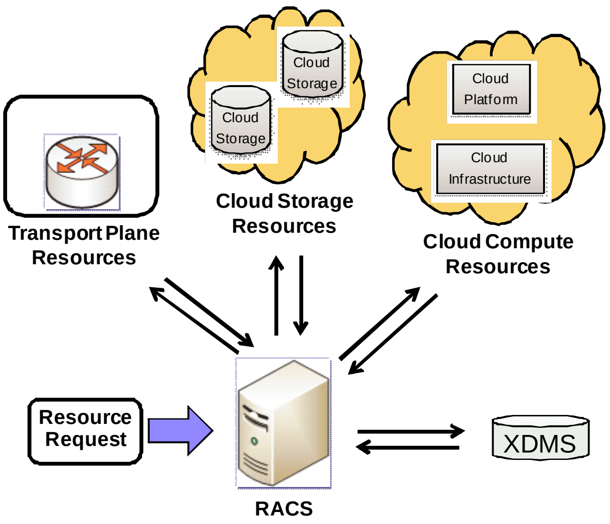

Fig. 7. RACS role players.

VI. CONCLUSION

Cloud computing is a new approach to delivering services over the Internet, and one from which NGN operators can benefit. This paper has presented grid and cloud technologies and discussed standardization gaps regarding interoperability between these technologies and the TISPAN NGN reference model. A unified architecture has been proposed that uses the PEEM to orchestrate IMS and Grid/Cloud services, and the RACS to expose grid and transport plane resources to any services. This unified architecture allows for the integration of IMS and Grid/Cloud services, which could results in new lucrative and innovative services. By viewing the Grid/Cloud domain as a diverse resource, the RACS can manage access through standardized, existing interfaces in a flexible and generic manner. This work has highlighting open areas regarding the integration of two important technologies and taken the first steps in proposing a unified NGN Grid/Cloud reference model.

REFERENCES

[1] ETSI ES 282 001 Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture Release 1

[2] ETSI TS 182 019 TISPAN Resource and Control Sub-system (RACS), functional architecture. ETSI TR 102 659-1, “GRID; Study of ICT Grid interoperability gaps; Part 1: Inventory of ICT Stakeholders”, V1.1.3 (2009).

[3] ETSI TR 102 659-2, “GRID; Study of ICT GRID interoperability gaps; Part 2: List of identified Gaps”, V1.1.3 (2009).

[4] Barbara Martini, Fabio Baroncelli, Piero Castoldi, Angelica Aprigliano, “Experimental validation of a service oriented network architecture applied to global Grid computing”, Proceedings of the First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS’05), IEEE, 2005.

[5] E. Gubbins, “How Telcos could conquer the Cloud”, *Telephony Online*, April 2009.

[6] OMA-ERELD-PEEM-V1_0-20090120-C, “Enabler Release Definition for PEEM”, Candidate Version 1.0 – 20 Jan 2009

[7] ETSI TISPAN, “ES 282 003 Resource and Admission Control Subsystem (RACS) Functional Architecture,” July 2008.

[8] http://aws.amazon.com/s3/

[9] http://aws.amazon.com/ec2/

[10] Geng Lin; Fu, D.; Jinzy Zhu; Dasmalchi, G.; “Cloud Computing: IT as a Service”, IT Professional, Volume 11, Issue 2, March-April 2009 Page(s):10 – 13.

[11] Leavitt, N.; “Is Cloud Computing Really Ready for Prime Time?”, Computer, Volume 42, Issue 1, Jan. 2009 Page(s):15 – 20.

[12] Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., and Brandic, I. “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility”, Future Generation Computer Systems, Volume 25, 6 (Jun. 2009), Pages: 599-616.

BIOGRAPHIES

Fabricio Carvalho de Gouveia (M.Sc) received his graduation in Electrical Engineering from the University Regional of Blumenau (Brazil) in 2000 and his M.Sc. in Telecommunications from the Federal University from Parana (Brazil) in 2003. He is employed as a Research Associate at the research center for Next Generation Network Infrastructure at FOKUS Fraunhofer Institute, where he is working towards his Ph.D in the Field of Next Generation Networks (NGN).

Richard Good received his B.Sc. (Hons) degree from the University of Cape Town in 2005. He is currently working towards his Ph.D. at the same institution. He is an active open source software contributor and has developed various open source IMS tools. His research interests include next generation resources management, service provisioning and QoS in heterogeneous networks.

Thomas Magedanz (PhD) is professor in the electrical engineering and computer sciences faculty at the Technische Universität Berlin, Germany. In addition, he is director of the “next generation network infrastructure” division of the Fraunhofer Institute FOKUS, which provides various testbeds and tools in the context of converging networks and open service delivery platforms. Since more than 20 years Prof. Magedanz is working in the convergence field of fixed and mobile telecommunications, the internet and information technologies. Under his leadership many service development platforms, toolkits and testbeds have been developed, such as the Grasshopper Mobil Agent platform, the OSA/Parlay Playground, the Open Source IMS Core System, and most recently the Open SOA Telco Playground. In the course of his research activities he published more than 200 technical papers/articles. In addition, Prof Magedanz is senior member of the IEEE, and editorial board member of several journals. In 2007, Prof. Magedanz joined the European FIRE (Future Internet Research and Experimental Facilities) Expert Group.