Cloud Computing and EPC / IMS Integration: New Value-Added Services on Demand

Fabricio Gouveia[1], Sebastian Wahle[1], Niklas Blum[1], Thomas Magedanz[2]
[1]Fraunhofer Institute FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany
{fabricio.gouveia | sebastian.wahle | niklas.blum}@fokus.fraunhofer.de
[2]Technische Universität Berlin, Franklinstraße 28-29, 10587, Berlin, Germany
tm@cs.tu-berlin.de

ABSTRACT

Cloud computing is a challenging and promising paradigm shift in the IT world. Examples of cloud services like Google Apps and Microsoft Azure represent a new paradigm for the provisioning of computer infrastructure, diminishing the costs for creating and maintaining software and hardware for users with infrastructural needs. Next Generation Networks (NGN) decouple services from the transport plane in order to provide flexibility with a common control architecture. This NGN approach offers faster integration of subsequent applications with the advantage of reusing the current deployed infrastructure. The integration of Cloud technology with Telecommunication infrastructures provides new revenue streams for Telecom operators through their converged fixed, mobile and data services and by offering "cloud computing data" services with NGN infrastructure features. This paper analyses interconnection scenarios for combining Cloud-based systems and an Evolved Packet Core (EPC) to offer new value-added applications in a unified architecture. The EPC and IP Multimedia Subsystem (IMS) are described as possible vehicles for the integration.

Keywords

Cloud Computing, IP Multimedia Subsystem, Next Generation Networks, Integration, Evolved Packet Core, Federation, Panlab

1. 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. Furthermore, this approach can lead to a fragmented end-user experience. For the deployment of new services, separate application servers need to be deployed, resulting in a linear increase of CAPEX and OPEX.

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. Therefore, OPEX costs are shared across the whole architecture. End-users have the benefit of enhanced experience through integrated services.

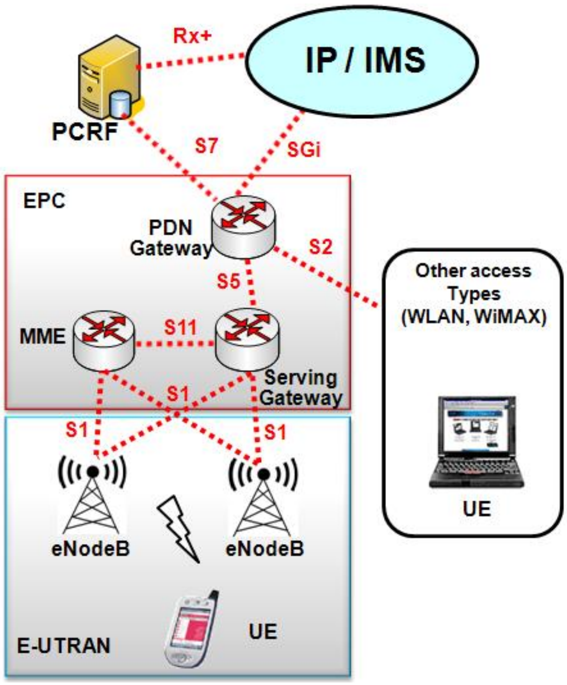

Figure 1: EPS reference architecture

The Third Generation Partnership Project (3GPP) specifies a solution to handle the heterogeneity of future networks, including QoS and charging aspects. It utilizes a policy-based approach based on the Policy and Charging Control (PCC) architecture, which is an integral component of the Evolved Packet System (EPS) [1]. The IMS forms part of the EPS (Fig. 1) as an IP service element and is the global standard for supporting multimedia services in IP networks. IMS offers increased reliability through efficient management of resources. It has been adopted by several other standardization bodies (ITU-T, ETSI TISPAN, 3GPP2, WiMAX Forum) for use in their NGN networks. Besides IMS, there is an Evolved Packet Core (EPC), which provides the necessary functions to control different access networks. Compared to other end-to-end architectures, EPS has high potential for mass-scale adoption, as it provides an upgrade to the

Mobimedia'09, September 7-9, 2009, London, UK.

current 3GPP 3G technology (UMTS), as well as a set of functionalities that comprise a complete management solution.

The ETSI Technical Committee (TC) GRID identified the gaps and overlaps that need to be addressed for GRID/TISPAN NGN [2],[3],[4] interconnection. Analogously, this paper analyses interconnection scenarios for combining Cloud-based systems and EPC to offer new value-added applications in a unified architecture. In particular, the advantages of such 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.

The remainder of the paper is organized as follows. Section 2 describes the basic concept of Cloud computing and its advantages. Section 3 identifies gaps in the TISPAN standardization for the integration of NGN and Grid environments. In Section 4 and 5, architectural scenarios for a unified architecture are discussed for a unified EPC and Cloud architecture. Section 6 concludes the paper and discusses future work directions.

2. CLOUD COMPUTING

The definitions of Cloud and Grid computing are closely related, open to interpretation and often misused. We need to distinguish between the two and know how they relate to one another. Grid computing is a concept based on the collaboration of distributed 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 [5]. It works as a parallel processing architecture with CPU resources shared across the network, to functions as a single supercomputer.

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. End-users of a Cloud service do not need to 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 [6].

The Cloud computing concept incorporates the delivery of Infrastructure (IaaS), Platform (PaaS) and Software as a Service (SaaS). Cloud computing intends to offer resources with the following characteristics: flexibility; abstracted resources featuring scalability, pay as you use model, reliability and 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. Cloud computing platforms have strong support for virtualization, dynamically composable services with Web Service interfaces, and strong support for creating 3rd party value-added services by building on computation, storage, and application services [6]. Nowadays, end user equipment is able to run applications within virtual machines efficiently. Virtual machines allow both the isolation of applications from the underlying hardware and other virtual machines, and the customization of the platform to suit the needs of the end-user. Providers can expose applications running within virtual machines or provide access to them as a service, which allows users to install and run their own applications.

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 computing clouds and storage clouds. The Amazon Elastic Computing Cloud (EC2) and Amazon Simple Storage Service (S3) are the pioneering applications in this sphere[7],[8]. We can cite here also the Microsoft Azure service platform [9], which represents another example of the Cloud computing paradigm.

3. 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 [10]. According to [11], working via large platforms owned by providers and shared by numerous users makes Cloud computing less expensive. Telco operations have many network components and software, which can be combined well with Cloud services. Each software component can be treated like a hosted application, which yields parallels in processing and performance [10]. 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.

In this context, IMS and NGN service architectures could fit comfortably in Cloud environments. 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 Cloud domain. Cloud service providers could also benefit from the unified management and control architecture and a bigger user base.

It is clear that a unified EPC/IMS and Cloud domain has advantages for all stakeholders including end-users (more usage and service options), network operators (different business models) and Cloud service providers (larger user community). However, there are architectural challenges for such a unified architecture. The ETSI Technical Report 102 659-2 [4] 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.

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 Cloud infrastructures.

Current gaps highlighted in the report include:

- Support of end-to-end service on Grid, 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 accepted architectural reference model.

To address these issues it is necessary to abstract the Cloud domain and view it as a diverse resource that can be controlled by the EPC components over standardized interfaces. Cloud resources like computing power, network and storage resources can be managed in a similar fashion to the NGN transport plane via the gateways and Mobility Management Element (MME). The question remains how to integrate these standardized interfaces into the 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.

4. INTEGRATION SCENARIOS

There are two primary architectural scenarios for combining NGN and Grid technologies [4]. These are: Applications implemented with Grid technology and a Grid architecture implemented as an NGN subsystem. Based on these concepts we define new possible scenarios for EPC / IMS integration. Basically, there is a gap on how to offer Cloud services to NGN users and vice-versa [4]. For instance, in order to provide a Grid-based video download service over NGN or an NGN-based video download service over Grid, the mechanisms used are different. NGN may require more information for authentication and authorization than Grid. Therefore it might not be possible to offer a Grid service that is realized by NGN.

For the first scenario the current NGN model will need to be extended in order to support Cloud services. Fig. 2 proposes the enhancement of the current EPC model to support cloud functionalities. This scenario is to provide the interconnection interfaces needed for the core infrastructure to be cloud-ready. In other words, the extension of the already standardized interfaces to cope with the principal cloud requirements should be available.

The main idea is to create a convergence platform that is capable of composing specific services to meet the needs of different groups of users, and therefore cover specific niches requiring the features from NGN with the flexibility and cost-efficiency of clouds.

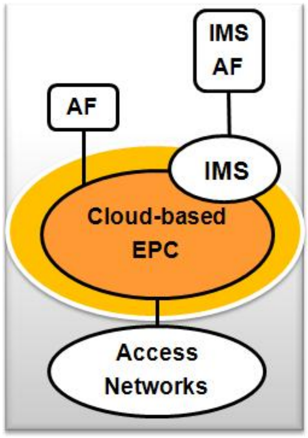

Figure 2: Implementing cloud functions in the EPC

In terms of network requirements, cloud applications are diverse. While it is not possible to define a generic Cloud 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. This generates two further scenarios considering the use of EPC and IMS, depicted in figures 3 and 4. Cloud enabled applications run within a Cloud Application Server, which has access via standardized interfaces to existing subsystems, e.g. ISC interface between IMS and Application Functions (AF).

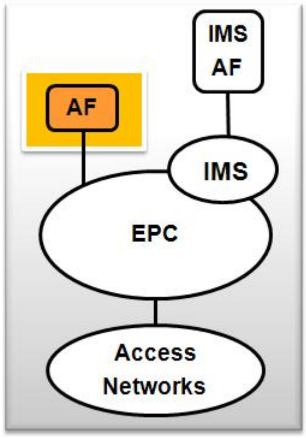

Figure 3: Develop and integrate a cloud-based application function

The ECP Cloud-enabled subsystem provides a service control architecture to support Cloud services through resource virtualization and a service interface. This facilitates three use cases: cloud services can be offered to cloud-enabled applications; cloud resources can be offered to end-user applications; dedicated cloud-enabled functions from other subsystems can interact with the enhanced EPC. The impact that this coordination has on the defined NGN reference points and the architecture as a whole is under investigation in TISPAN [4].

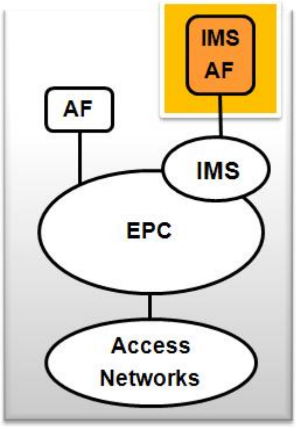

Figure 4: Develop and integrate a cloud-based

IMS application function

Provisioning of Cloud services with the functional features of NGN requires a framework for interworking with all players involved in the delivery process. 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 and EPC, the scenarios from figures 2, 3 and 4 can be implemented based on the principles of OMA Policy Evaluation Enforcement and Management (PEEM) [12] and the EPC to control transport resources.

5. PEEM ENABLER FOR CONTROLLING CLOUD SERVICES

New services developed by 3rd Party service providers (e.g., Web Services, Cloud 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. However, such personal information and network resources have to be protected. A central point that enforces service composition, orchestration, delegation and also controls access to enablers, applications and network resources is needed.

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 [12].

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 innovative services. Fig. 5 depicts the context of PEEM to offer Third Party Service Control in 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 pre-defined policies with each incoming request and enforces the decision result.

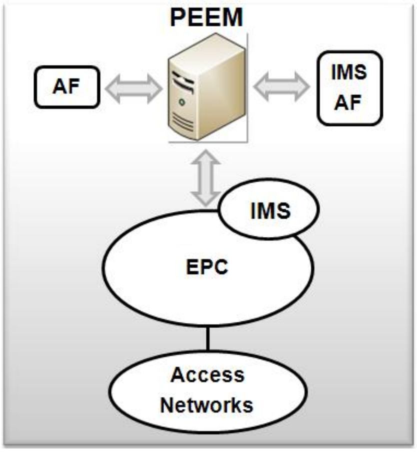

Figure 5: The use of PEEM to offer third party service control to the NGN comain

These features motivate the use of the PEEM to control service requests from Clouds 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.

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 Application Programming Interfaces (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. The usage of resources like computing power and storage in cloud services is analogous to transport plane resources, like bandwidth, that are consumed by IMS sessions. Requests from cloud 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 Policy and Charging Control (PCC) is defined as a functional element within the NGN to expose resource management functions to applications [13]. The main objective of the PCC is to interconnect the signaling plane with the data plane, providing access control, resource control and QoS control. The PCC is therefore a complex layer where the higher level of abstraction achieved in the core network of IMS connects to the different access technologies in a seamless way. Since the first IMS standards there have been three different specifications of the functionalities in charge of policing, QoS control and charging. Applications compose authorization requests; the PCC 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 [13].

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

The use of PEEM and EPC in the described integration scenarios is envisaged as a first step towards cloud support in NGN environments. Advanced management processes shall simplify integration and distributed deployment by means of a resource federation framework developed in Panlab [14], [15]. Panlab is based on a federation of distributed facilities that are interconnected, providing access to different platforms, networks, and services for broad testing and experimentation purposes. In this regard, the central federation control tool *Teagle* will play an important role in managing the described scenarios [16].

6. CONCLUSION

Cloud computing is a new paradigm delivering IT services and infrastructure as computing utilities. NGN operators can benefit from this new paradigm and extend the NGN infrastructure in order to support clouds and generate more revenues. This paper has presented briefly the concept of Cloud computing and discussed standardization gaps regarding interoperability between this technology and the TISPAN NGN reference model. Integration scenarios have been proposed for integrating Cloud and NGN. The approach uses the PEEM to orchestrate IMS and Cloud services, with the use of the EPC and PCC architectures to control transport plane resources. This unified architecture allows for the integration of IMS and Cloud services, which could result in new lucrative and innovative services. The solution offers a more dynamic infrastructure in the context of NGN because it creates an adaptive infrastructure, improving flexibility and reducing operating costs. By viewing the Cloud domain as a diverse resource, the EPC and PCC can manage access through standardized, existing interfaces in a flexible and generic manner.

7. AKNOWLEDGEMENTS

We would like to thank Richard Good for his valuable contributions to the analysis of NGN and Cloud/Grid integration in the standardization documents. This work investigates first scenarios for the integration of cloud networks with NGN and are envisaged to be applied later in the experimental facility scenarios from the Panlab project.

8. REFERENCES

[1] Lescuyer, P., & Lucidarme, T. (2008). Evolved Packet System. The LTE and SAE Evilution of 3G UMTS. John Wiley and Sons, Ltd. ISBN: 978-0-470-05976-0, pp. 350.

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

[3] 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).

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

[5] Martini, B., Baroncelli, F., Castoldi, P., Aprigliano, A., "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.

[6] 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.

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

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

[9] Microsoft, "Azure service platform," available at http://www.microsoft.com/azure/data.mspx, 2008.

[10] E. Gubbins, "How Telcos could conquer the Cloud", Telephony Online, April 2009.

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

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

[13] 3GPP TS 23.203, "Policy and charging control architecture", V9.0.0, March 2009.

[14] Campowsky, K.; Gavras, A.; Harjoc, B.; Magedanz, T. and Wahle, S. "Pan-European Testbed and Experimental Facility Federation – Architecture Refinement and Implementation" Inderscience International Journal of Communication Networks and Distributed Systems (IJCNDS), Special Issue: Recent Advances in Test-bed Driven Networking Research. Accepted for publication in 2010.

[15] The Panlab and PII project website: http://www.panlab.org/

[16] Magedanz, T.; Schreiner, F. and Wahle, S. "Service-Oriented Testbed Infrastructures and Cross-Domain Federation for Future Internet Research" In 2009 IFIP/IEEE International Symposium on Integrated Network Management Proceedings, New York, USA, June 2009. IEEE.