MOBILE CLOUD COMPUTING BASED ON SERVICE ORIENTED ARCHITECTURE: EMBRACING NETWORK AS A SERVICE FOR 3RD PARTY APPLICATION SERVICE PROVIDERS

Michael Andres Feliu Gutierrez, Neco Ventura

Communication Research Group
Broadband Networks and Applications
Department of Electrical Engineering
University of Cape Town, Rondebosch, 7700
Email: {mfeliu, neco}@crg.ee.uct.ac.za

ABSTRACT

The recent emergence of Cloud Computing in the IT world has opened doors for new revenue streams and business models. The movement towards a service-oriented architecture and an all-IP based communication system has led to the IP Multimedia Subsystem (IMS) being accepted as the Next Generation Networks (NGN) service control-provisioning platform. The Telco 2.0 domain can benefit from service enhancement and Web 2.0 technologies such as Cloud Computing. This can be achieved by opening gateways and APIs guarding rich underlying network resources (e.g. location) to 3rd Party ASPs, thus adopting the delivery of Network as a Service (NaaS). Telco services are faced with opportunities to make use of powerful computational power and storage services offered by cloud environments to accelerate business-processing speeds.

Keywords— Cloud Computing, NGN/IMS, *aaS, Web 2.0, OMA PEEM enabler

1. INTRODUCTION

Over the years the telecommunication industry's main revenue stream has come from voice applications, in recent years the mobile operators have been seeking for new revenue streams and business models.

There have been debates on whether mobile and broadband communication should be provided as a commodity. It is stated in [1] that the IT and Telecommunication industries are undergoing transformations of their infrastructure with the intentions to deliver services in a way similar to traditional utilities like electricity, water and gas.

The Australian government began a project [2] that will last 10 years with the intentions to deliver computing as a 5th utility. Computing is brought to your household and users consume in a "pay as go" basis.

On the other hand Cisco's Global Mobile Data Traffic Forecast [3] has predicted an overall growth of 6.3 Petabytes per month by 2015. This excess growth of data traffic as can be observed in Figure 1 below [3] is mainly due to the mobile world with user equipment such as smartphones and laptops.

Mobile video and mobile data will contribute to almost 90% of the mobile traffic growth, with video due to its high bit rates contributing 66%. This tremendous increase in data flow brings challenges to the mobile operators (e.g. how to cope with gigantic storage and enormous amount of processing).

.

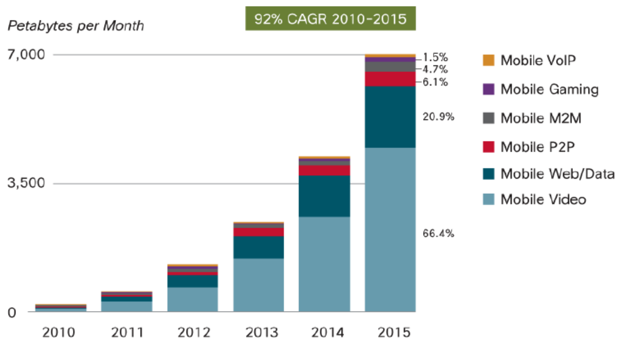

Figure 1. Showing the major contributors to the mobile traffic growth.

It can be seen that today there is an increasing demand for multimedia services and sophisticated applications, which in turn brings together internet applications with telecommunication applications. This gives the telecommunication operators great opportunities to benefit from the emerging Web 2.0 technologies. Simple protocols like HTTP, HTML description language and SOAP/REST web-based services have driven the developers community in creating sophisticated content-aware applications.

The Telco industry has to adapt to the customer's needs by upgrading their systems, installing new hardware/software and integrating new technologies into their existing infrastructure. Keeping in mind that one of the key aspects of NGN telecommunication platforms is the reuse of its

existing infrastructure for new market driven applications through dedicated application enablers, thus sharing operational expenditure across the whole architecture.
This paper discusses a service creation mechanism with strong support for creating 3rd Party ASPs value-added services by building on computation and storage in a cloud environment and application services in NGN/IMS environment. Section II describes cloud computing in both worlds with the different business models and service delivery adaption. Section III discusses a NGN service-oriented architecture with focus on the IMS Application Servers (A.S.) for service delivery. Section IV proposes a possible integration scenario for a mixed Cloud/IMS environment. Section V presents some results obtained during the integration analysis. Section VI concludes this paper and states some final remarks.

2. MOBILE CLOUD COMPUTING

In the last decade new technologies have emerged, one of particular interest and gaining popularity in the Web 2.0 domain is cloud computing. The term Web 2.0 was first used by Tim O'Reilly [16] to describe a fast-growing set of web-based applications.

Cloud Computing is a concept that is currently enjoying much hype and its exact definitions is open to interpretation. Generally speaking it refers to the sharing of some kind of hosted service or application. It can be defined as a paradigm shift in the IT world where computing resources (e.g. processing power, storage) are moved away from the user's personal computer or application server to a cloud of powerful computers [1, 4].

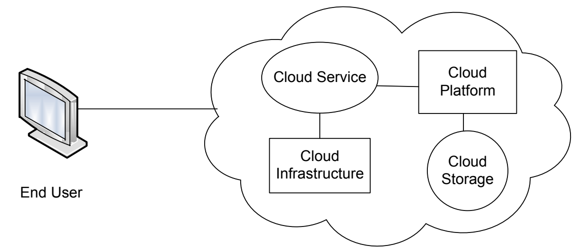

Figure 2. Showing the benefits of cloud computing

Cloud computing [6] can supply transparent and on-demand access to applications served over the Internet in a dynamic and scalable manner. It adapts the delivery of various classes of services.

2.1. Software as a Service (SaaS)

Software as a Service can be viewed as an improved version of the Application Service Provider (ASP) model in which a service provider host software applications over the network. The service is hosted/ located on a cloud environment where it can be accessed on demand by a UE anytime anywhere. SaaS can be categorized under Business-oriented SaaS (e.g. SalesForce, Amazon S3) and Consumer-oriented SaaS (storage, Facebook).

2.2. Platform as a Service (PaaS)

Platform as a Service are hosted application environments where software developers/ 3rd party ASP who want to focus primarily on the service development cycle and the orchestration of new services, can bypass the capital expenditure that would otherwise be needed for the deployment of the underlying infrastructure. PaaS examples include the Google App Engine where developers can enhance an existing SaaS by providing mash-ups or simply develop new web-based applications.

2.3. Infrastructure as a Service (IaaS)

Infrastructure as a Service can be described as utility computing data centers that make use of cluster virtualization technology to provide powerful and flexible computing resources. IaaS provides on demand resources such as parallel computing power to process large amount of data and virtual storage to store gigantic data. Examples [5] include Microsoft Windows Azure and open source EUCALYPTUS and Nimbus.

2.4. Network as a Service (NaaS)

Network as a Service is a term used in the Mobile world [8] referring to Telco operators opening up network APIs to expose network capabilities (e.g. presence, location, billing and charging). These network resources are exposed in a standard and flexible manner to 3rd Party ASPs where they can mash-up Telco services with IT services and vise versa.

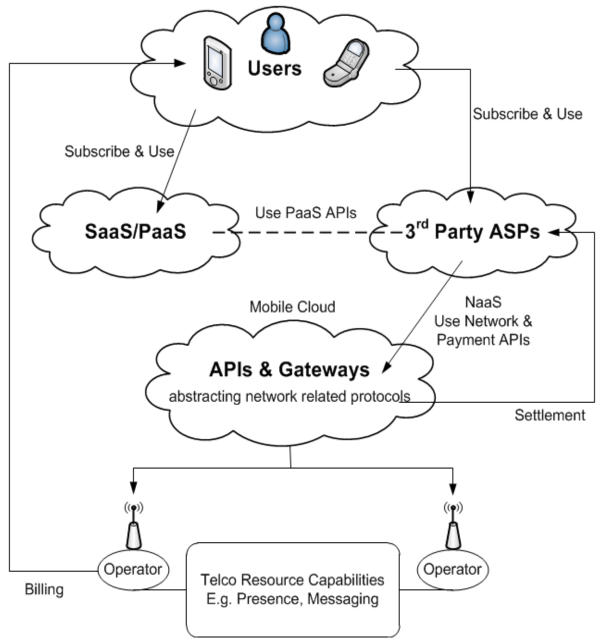

Figure 3. Showing the concept of NaaS

Telco operators can benefit from the delivery of network resources as services and give rise to the Telco 2.0 two-sided business model. The advantage would not only be revenue from their users but also from 3rd Party ASPs who use network resources to build applications and may adapt a pay as you go basis depending on the SLA.

The cloud architectures extend to the end user by allowing web browsers and software applications residing inside/ outside the cloud domain to access cloud resources. The two major forms of cloud resources are computing clouds and storage clouds.

3. SERVICE ORIENTED ARCHITECTURE

The European Telecommunication Standards Institute (ETSI) defined the Internet Protocol Multimedia Subsystem (IMS) in [7] as an overlay network where the service logic has been stripped from the transport and network protocols to facilitate and standardize the service delivery mechanism.

NGN supports end-to-end services, where the QoS is independent of the underlying networking technology. Interoperability among multiple network service providers is facilitated with such architecture. A SOA thus facilitates the reuse of software and enables the creation of composite services with intent to write-once, run and sell anywhere goal.

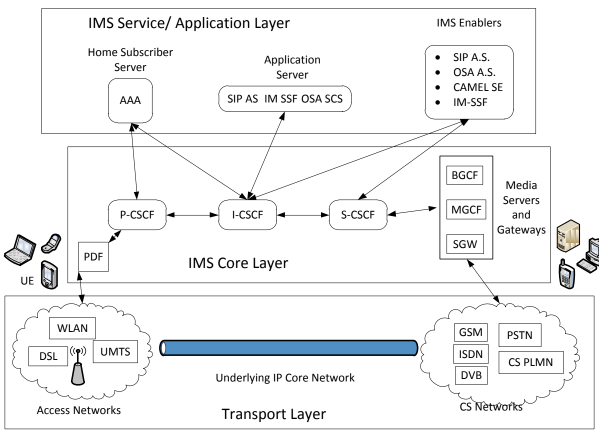

Figure 4. Showing the NGN/IMS reference architecture

3.1. Application Layer

The service logic is implemented in the A.S and is thus the service relevant part of the IMS. Programmer's are not limited to their favorite programming language, common means of implementation include C++, SIP applications, Java. A.S are considered value-added services, which can be used to enhance or mash-up already existing services. With the use of SIP application on top of SIP servers allows four mode of operation.

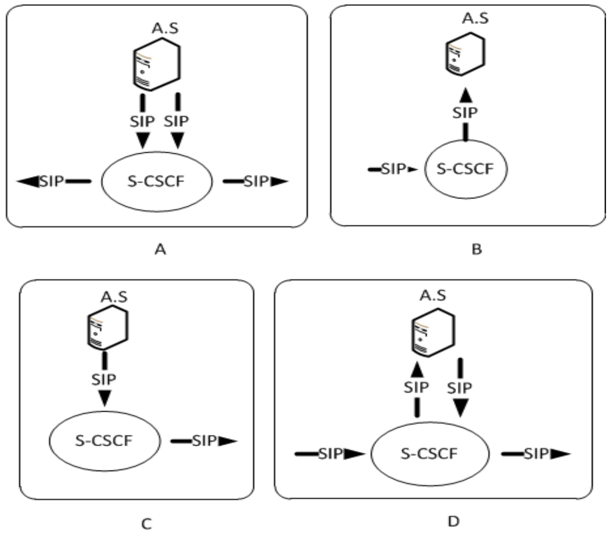

Figure 5. Showing A.S mode of operation
A – A.S acting as a Back-to-Back UA
B - A.S acting as a terminating UA
C - A.S. acting as an originating UA
D - A.S. acting as a SIP proxy

The programming flexibility and the movement towards a service-oriented programming model can be illustrated in Figure 5 which shows the different roles an A.S. can have and the different outcomes in terms of SIP signaling.

The IMS was developed as the next generation core architecture for converged voice and data services, which provides a common platform for different access technologies like Wi-Fi, Wi-Max, and DSL and aims to supply an open, standard based network that grants a wide range of multimedia services. This motivated the use of the IMS for this project in order to facilitate the deployment of services which not only uses IMS related capabilities (e.g. IM, presence) but also resources (processing, storage) from other networks residing outside the operator's network.

3.2. IMS-related Entities and Functionalities

There are 3 main Call Session Control Functions, each with its own special task, all working together to handle SIP requests, registration and session establishment.

- Proxy-CSCF is the 1st contact/entry point into the IMS environment from the client UE, it's responsible for compressing data, interacting with policy and charging rules and providing security
- Interrogating-CSCF is an intermediate point whose function is to provide topology hiding by obtaining the next signaling hop, either an A.S or S-CSCF
- Serving-CSCF is the focal/anchor point in the IMS as it is responsible for handling registration, storing service profiles and maintaining session states

The IMS entities communicate via two main interfaces forming the central session control protocols defined by IETF.

1. SIP - Session Initiation Protocol is an end-to-end client-server session signaling protocol defined by the Internet Engineering Task Force (IETF), used for session creation and termination, as well as for providing services such as presence
2. DIAMETER is used to perform AAA operations Authentication, Authorization and Accounting as well as interfaces with the HSS database to download and upload user profiles

For clarity sake only two protocols were mentioned but it is important to note that the IMS entities communicates with a vast number of well defined interfaces [11] that are out of the scope of this paper.

The Home Subscriber Server (HSS) is the main data storage database for all subscribers and service-related data (e.g. memory and bandwidth requirements). Popular databases include MySQL where registration information and service information (e.g. target resources, memory requirement) can be easily stored and obtained.

4. NGN/IMS AND CLOUD INTEGRATION SCENARIO

NGN applications running on smartphones and netbooks will operate almost entirely off mobile cloud based services. This will be achieved with the promised downlink and uplink speeds of 4G. Over time, the mobile cloud may do for the mobile devices what SaaS has done for the desktop, connecting people, services and digital data and in a flexible and standardized manner.

In order to integrate the two technologies separate standalone environment had to be deployed and tested. The environments are limited to hardware limitation.

According to the ETSI [13] there are a number of possible integration scenarios that can be achieved, the benefits and impacts of these scenarios on the NGN platform are still in its infancy and it's an open area of research currently gaining hype in both worlds.

4.1. Cloud Computing Environment

An IaaS environment was deployed using the open-source [10] Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems (EUCALYPTUS). It is a simple, flexible and modular cloud computing framework that uses computational and storage infrastructure. Resource access latency is greatly reduced by using Xen virtualization technology; it supports Virtual Machines that run on top of the Xen [10] hypervisor.

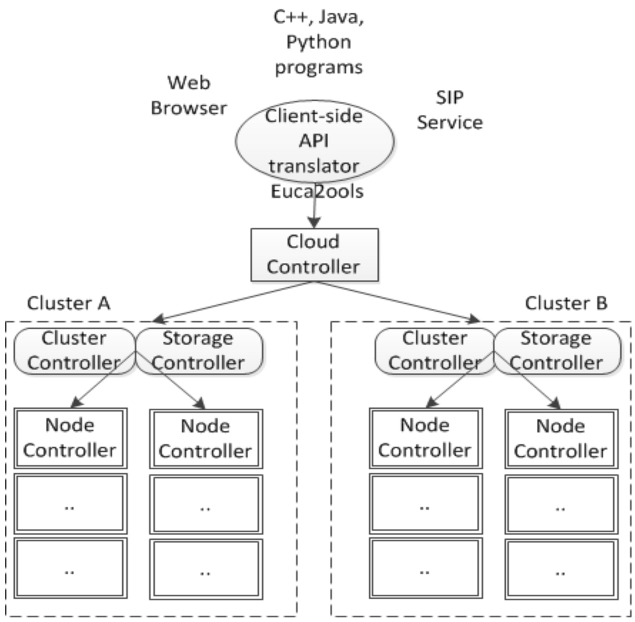

Figure 6. Showing eucalyptus architecture

The eucalyptus platform [7] is composed of the following main components

- Cloud Controller – is responsible for the management of the virtualized resources (e.g. storage, CPU instances) that are offered on demand. It is important to note that this controls the provisioning of these resources in the cloud domain, but how best to utilize these resources is up to the software developer/ 3rd Party ASPs
- Cluster Controller – is responsible for the controlling virtual machines/servers running on nodes
- Storage Controller – provides block-level storage that can be accessed on demand and can be dynamically encapsulated by virtual machines instances
- Node Controller – control underlying VM or instances by the use of a hypervisor
  - Execution, Management
  - Termination, Inspection

Xen [10] is the hypervisor of choice for controlling instances due to its compatibility with a Linux OS, where Eucalyptus is deployed.

In fact IaaS platform like Amazon place server clusters across the globe in order to improve the responsiveness, locality and redundancy of the content it hosts for users. However their price ranges are out of the reach of most and often require lengthy contract and a large usage commitment. For this reason a simple open source small-scale IaaS environment was deployed to take advantage of the processing and storage they offer.

NGN applications can be considered to be a mash-up of IT services using computational and storage resources interacting with Telco resources capabilities (e.g. presence, messaging) in a flexible manner. These applications can be computationally demanding and the UE is limited to its inbuilt processing power and storage capacity. On demand access to storage and cluster computation can be made available through the IaaS environment where services deployed can access the resources in the cloud environment.

4.2. IMS Environment

The IMS platform is based on the FOKUS Open IMS, which has been deployed as a testbed for multimedia services at the UCT CoE. The IMS core was deployed on a Linux-Ubuntu kernel running on top of an Intel core2 duo CPU @ 2.66GHz and with 4GB RAM. All service related information is stored in a lightweight HSS (FHoSS), which relies on MySQL; the database is thus populated using SQL scripts included in the IMS package.

The UCT open IMS Client [15] is a well-defined IMS end user client that communicates with the IMS entities and ultimately provides services to the customer in an abstract way. The IMS client can initiate an instant messaging services using SIP MESSAGE method, alternatively the Open Mobile Alliance (OMA) IM enabler [11] can be used which adds value to the already existing IM service by adding for example buddy list or possibly store messages.

IMS resources are expressed and interconnected via Uniform Resource Identifiers (URI), which is in the form of *sip:R1.resources@anIMSprovider.com*.

4.3. PEEM-based Serving Gateway (PSG)

The Serving Gateway sits at the border between the IMS and the Cloud environments. It acts as an intermediate logical entity, intercepting request from both worlds. It allows the exposure of IMS resources to 3rd Party ASP by acting as a SIP to HTTP protocol translator. Conversion is made simple due to the fact that SIP [12] borrows and adapts methods from HTTP [17]. It exposes web services and SIP applications to the managed, reliable and flexible IMS platform.

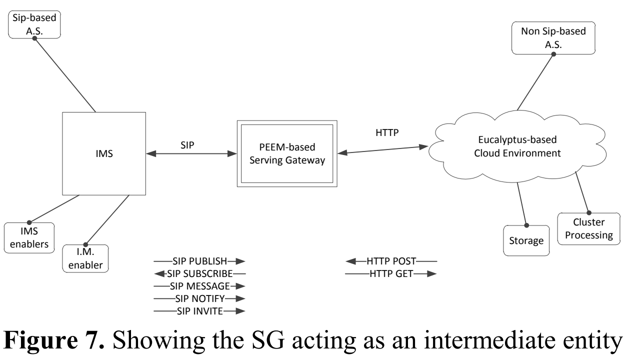

Figure 7. Showing the SG acting as an intermediate entity

The Open Mobile Alliance (OMA) standardization body has defined the Policy Enforcement Evaluation Management (PEEM) enabler. The OMA defines a number of usage patterns for the PEEM enabler, which can be used for vast number of applications. It is important to note that PEEM proxy usage pattern [14] is the most relevant to this project as it can be used to intercept service requests from a foreign domain and from other service requester generated by 3rd Party ASP. Policies (rules) can be evaluated and enforced depending on the SLA. The PEEM can be used to expose network capabilities like presence and IM to the cloud domain. The Serving Gateway contains among others, a ContextHandler function where HTTP and SIP request are sent. It is here where rules are enforced or evaluated.

5. RESULTS

A change of media scenario between an IMS service and cloud service residing in the cloud domain is favorable. A simple use case was built around this idea to test the seamless switch between an IM session and a write/read session using storage instances provided in the cloud environment.

The size of a file usually determines the processing time required to store that file. Multithreading and parallel architecture achieved in a cloud environment allows the processing task to be sub-divided or split among independent threads or instances running concurrently. Depending on the SLA a client is allowed to use a certain amount of processing power usually on-demand access adapts pay as you go, thus you pay for exactly how much processing power you need.

```
VNET_DHCPDAEMON=/usr/sbin/dhcpd
VNET_PRIVINTERFACE=eth1
VNET_MODE=STATIC
VNET_SUBNET=137.158.125.0
VNET_NETMASK=255.255.255.0
VNET_BROADCAST=137.158.125.20
VNET_ROUTER=137.158.125.1
VNET_DNS=137.158.125.56
VNET_MACMAP=00:1D:92:60:2F:6A=137.158
.125.25100:1D:92:60:2F:6B=137.158.125
.25200:1D:92:60:2F:6C=137.158.125.253
```

Figure 8. Showing hypervisor snippet while testing instance availability in the cloud infrastructure deployed.

The snippet shows three IP addresses that are available for instances. A 3rd Party ASP can use 137.158.125.251-3 to obtain on demand access to three cluster processing instances.

The diagram below illustrates the adoption of two SLAs and shows the processing time where two clients with the same file size have different amount of processing power. Thus one client uses one instance of processing while the other uses three instances to execute the same task.

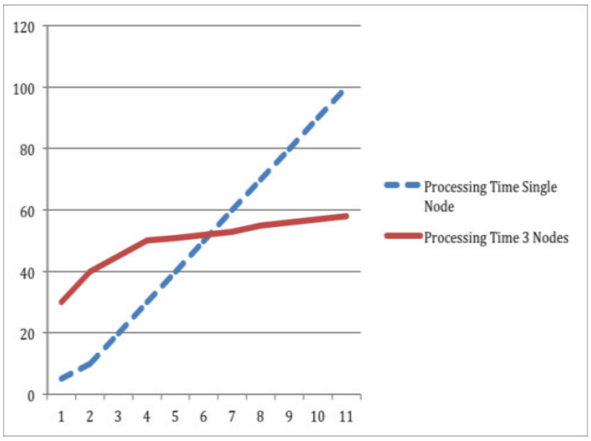

Figure 9. Showing the concept of on demand processing power depending on two different SLA

Consider an IM session where Alice generates a message consisting of text and fills the request-URI with the address of Bob. The IMS infrastructure (P/I/S-CSCF) forwards the message to the recipient. Once bob receives the message, he replies with a 200 OK message, as it can be observed in Figure 11 acknowledging the delivery. It is important to note that each instant message is an independent transaction and is not related to the previous requests. The protocol responsible for conveying messages within an instant messaging (IM) session is called Message Session Relay Protocol [18], which sits on top TCP and allows messages of arbitrary size, due to messages being sent in small chunks.

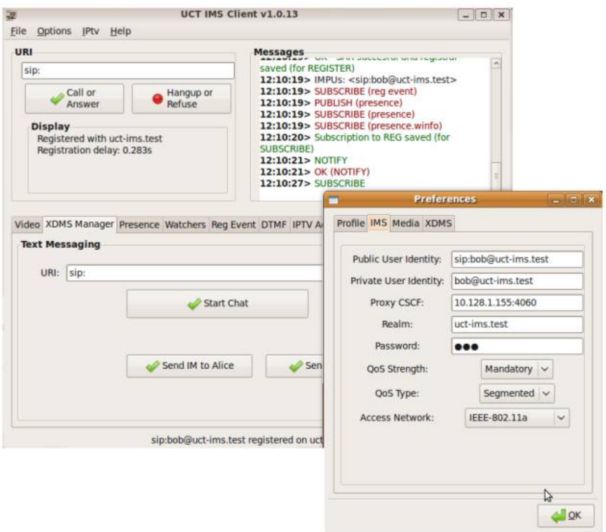

Figure 10. Showing the UCT IMS client instant messaging registration and configuration

To discover IMS resources a SIP SUBSCRIBE is sent by the SG and to discover cloud resources an HTTP GET can be sent by the SG. Alice later decides to access on demand storage from the cloud and give rights to bob to access her file (e.g. photo, video, letter). The SG is responsible for delegating services between the two worlds.

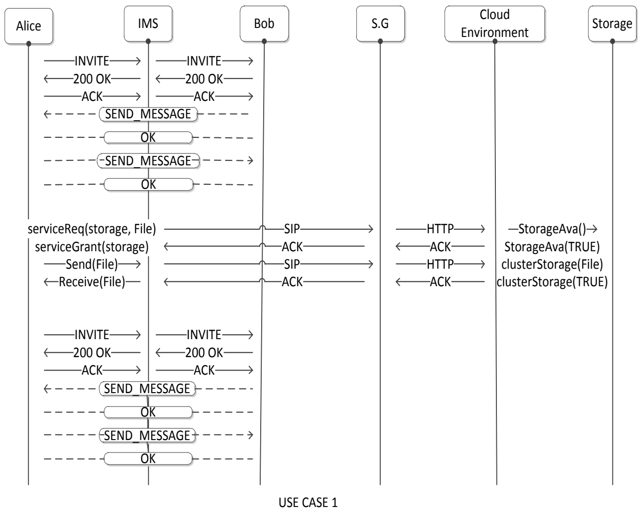

Figure 11. Showing signaling diagram for a simple use case

This use-case illustrates how we seamlessly expose the IMS presence and instant messaging done within the operator domain to access a storage instances provided by an external party on the Cloud infrastructure. This enables the provision of a consistent and efficient user experience, wherever the resource is stored and is network connection independent.

Currently many telecom service systems could be running in the "cloud", which make the services easy to manage, update and use. It can also make full use of the powerful processing ability and storage.

6. CONCLUSIONS AND FUTURE WORK

In the Web 2.0 domain, cloud computing has adapted the delivery of IT service and infrastructure as computing utilities. Telco operators are expecting that cloud-enabled application can improve their internal network operation in terms of allowing 3rd Party ASPs to enhance their existing services or develop new innovative services by mashing-up the IT and Telecom services in a flexible and attractive manner.

This paper briefly presents a possible scenario in which cloud application (on demand storage access) can be merged with a Telco application (instant messaging using presence).

Resource allocation algorithms are responsible for delegating resources to service request depending on request need, availability and allocation, thus providing as much resources as possible while preventing deadlock stages.

A system in practice consists of a finite number of resources that are shared among a number of competing processes. Our next step is to develop a resource-allocation algorithm like the Banker's Algorithm for multiple instances of each resource type including and not limited to disk storage, processing instances, and memory allocation. The aim is to increase system utilization when resources limitations are faced due to perhaps peak time usage or abnormal increase in resource demand or simply service attractiveness.

REFERENCES

[1] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility" Future Generation Computer Systems Journal Elsevier 2010

[2] The Broadband Advisory Group's Report to Government. "Australia's Broadband Connectivity" Summary Document

[3] Cisco White Paper, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010 - 2015" February 1, 2011

[4] John W.Rittinghouse, James F.Ransome. "Cloud Computing Implementation, Management and Security" CRC Press Taylor and Francis Group, 2010

[5] Asma BEN LETAIFA, Amel HAJI, Maha JEBALIA, Sami TABBANE, "State of the Art and Research Challenges of new services architecture technologies: Virtualization, SOA and Cloud Computing" International Journal of Grid and Distributed Computing Vol. 3, No. 4, December, 2010

[6] Vânia Gonçalves, Pieter Ballon, "Adding value to the network: Mobile operators' experiments with Software-as-a-Service and Platform-as-a-Service models" Telematics and Informatics Journals, Elsevier 2010

[7] ETSI ES 282 007, "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IP Multimedia Subsystem (IMS); Functional architecture" V2.0.0 (2008-03)

[8] AEPONA White Paper "Network as a Service and Mobile Cloud Computing." February 2010.

[9] Fabricio Gouveia, Sebastian Wahle, Niklas Blum, Thomas Magedanz. "Cloud Computing and EPC/IMS Integration: New Value-Added Service on Demand" Mobimedia '09**:** Proceedings of the 5th International ICST Mobile Multimedia Communications Conference, September 2009.

[10] Daniel Nurmi, Rich Wolski, Chris Grzegorczyk Graziano Obertelli, Sunil Soman, Lamia Youseff, Dmitrii Zagorodnov, "The Eucalyptus Open-source Cloud-computing System" University of California, Santa Barbara

[11] Miikka Poikselka, Georg Mayer. "The IMS – IP Multimedia Concepts and Services" Wiley Book 2009

[12] IETF RFC 3261, "SIP: Session Initiation Protocol". 2002

[13] ETSI TR 102 767, "Grid Services and Telecom Networks; Architectural Options" V1.1.1 (2009-02)

[14] Open Mobile Alliance, "Policy Evaluation, Enforcement and Management Architecture" V 1.0.0 Aug 2008.

[15] David Waiting, Richard Good, Richard Spiers and Neco Ventura, "The UCT IMS Client" Testbeds and Research Infrastructure for the Development of Networks & Communities and Workshops, TridentCom 2009

[16] Tim O'Reilly, "What is Web 2.0" O'Reilly Media Release September 2005.

[17] IETF RFC, 2616, "Hypertext Transfer Protocol - HTTP" 1999.

[18] IETF RFC 4975, "Message Session Relay Protocol - MSRP" September 2007.