

Architecture and Key Issues of IMS-based Cloud Computing

Wei Zhang, Weimin Lei, Xiao Chen, Shaowei Liu

Institute of Communication and Information System
College of Information Science and Engineering, Northeastern University
Shenyang, China
zhangwei1@ise.neu.edu.cn

Abstract—Cloud computing is changing the way of developing, deploying and managing applications. However, as typical Internet-based applications, cloud computing services lack carrier-grade signaling control mechanism and cannot guaranty Quality of Service (QoS), which have actually become technical barriers for telecom operator to provide commercial public cloud services. On the other hand, as the core signaling architecture of Next Generation Networking (NGN), IP Multimedia Subsystem (IMS) is facing the problem of the lack of innovative value-added services. This paper presents an architecture to support cloud computing services over IMS. In the proposed architecture, cloud services are regarded as the general IMS applications and then cloud clients are allowed to access cloud services under the control of Session Initiation Protocol (SIP) signaling and QoS mechanism of IMS. This paper introduces architecture overview and cloud service relevant functional components, and mainly discusses several key issues including cloud notification mechanism, QoS and charging control of IMS-based cloud computing services.

Keywords—Cloud Computing; IP Multimedia Subsystem; Session Initiation Protocol; Quality of Service; Charging

I. INTRODUCTION

Cloud computing represents the major technology trend in IT in the next decade. Cloud computing provides the powerful computing power and nearly unlimited storage space by sharing hardware and software resources, which facilitates the rapid deployment of a variety of innovative services in a cost effective manner. However, the technical progress and popularity of cloud computing haven't meet people's expectations. In addition to lack of technology maturity, the following factors are the key issues restricting to its wider application.

First, cloud computing lacks a complete signaling control mechanism. Almost all of the cloud platforms extend to the client via web service interfaces. However, the web-based access mode cannot effectively support finer granularity of service access control, various pricing models and other mechanisms. Therefore, it is difficult for telecom operators to provide commercial public cloud services.

Second, as typical Internet-based applications, cloud computing services cannot guaranty Quality of Service (QoS) for cloud clients. No matter what type of cloud services provided, cloud computing systems put forward a high demand for network bandwidth. According to the cloud computing's definition provided by National Institute of

Standards and Technology (NIST), one of key characteristics of cloud computing is broad network access [1]. However, Internet's best effort strategy, which is only appropriate for the early internet services, is difficult to guaranty QoS for cloud services, thus diminishes the quality of user experience.

Finally, protocols and standards of cloud computing service interface have not yet been universally established. According to their specific application requirements and existing technology advantages, the enterprises research and design their own cloud computing systems and offer their proprietary interfaces. At present, most of public cloud platforms (Amazon EC2[2], etc.) are based on proprietary service management mechanisms which complicate interoperability between clouds. Cloud clients have to deal with specific and daunting tasks for each cloud platform and cannot change among different cloud platforms straightforwardly, which can lead to a fragmented end-user experience.

The IP multimedia subsystem (IMS) [3], initially introduced by 3rd Generation Partnership Project (3GPP), now is evolved to be the core signaling architecture of Next Generation Networking (NGN) for multimedia services and has been widely deployed by telecom operators throughout the world. The key feature of IMS is the ability to provide standard signaling control and negotiable QoS for IP services. However, IMS has not been used on a wide scale due to the lack of innovative services yet.

This paper proposes an architecture to integrate cloud computing services into IMS. IMS provides an open and standard service platform and cloud services are regarded as the general IMS applications. Combined with the characteristics of IMS and cloud computing, such integration will bring both of them an explosive growth. Based on cloud computing where most applications run in the cloud, user terminals only need capacity such as internet access, audio/video decoding and interactive processing. Thus, cloud computing technology will lead to rapid developments of IMS value-added services.

Meanwhile, IMS provides the most significant opportunities for cloud computing.

- Open and standardized signaling control
- IMS provides perfect signaling control mechanisms which can implement fine-grained service access control such as digital copyright protection, charging, security.
- Optional, negotiable and differential QoS control

Interacting with the network elements that transport application flows, IMS can offer negotiable QoS for IP multimedia sessions not only at the time of establishment, but also during the session. And IMS can provide different qualities depending on user profile, location, access network and devices, etc.

- Service Reusability

The existing IMS services including presence, group management, authentication, and capability negotiation could be exposed via standardized interfaces to the cloud services. In addition, some basic cloud services should also be made available to other complicated cloud services.

- Uniform standardized cloud computing interfaces

To be deployed based on IMS architecture, all cloud services should follow the uniformly standard interfaces, which will promote standardization work of cloud computing.

The remainder of the paper is organized as follows. Section II describes the background and related work in IMS and cloud computing. Section III presents the architecture and functional components of IMS-based cloud computing. Section IV discusses several key issues including cloud notification service, QoS and charging control. Section V concludes the paper and discusses future work.

II. BACKGROUND AND RELATED WORK

IMS was introduced in the 3GPP architecture release 5 to support IP multimedia services including voice, video, audio, and text transmissions. The collaboration between IETF, 3GPP, and 3GPP2 is essential to the development of IMS for next-generation networks [4]. Now IMS is evolved to support IP multimedia services in the integrated all-IP networks. IMS enables faster integration of subsequent value-added services that take advantage of the current deployed infrastructure and all the core services it provides (presence information, session control, QoS, charging, etc). Currently, available IMS business services are still in their infancy. To obtain openness and interoperability, IMS exploits open standard IETF protocols wherever possible. In particular, IMS uses the Session Initiation Protocol (SIP) [5] for signaling and session management. SIP is an application layer signaling protocol for establishing and tearing down multimedia sessions such as Internet conferencing, telephony, presence, event notification and instant messaging.

Up to present, cloud computing has been attracting more and more attentions from both academia and industry. Cloud computing providers can offer their services at low cost due to the economic scales [6]. There are more and more public clouds and as well private clouds built by institutions and companies. Unfortunately, these clouds are often incompatible from each other, which increases the complication of porting applications between clouds as well as introduces potential vendor lock-in. Most recently, many efforts have been made to develop open standards for cloud computing, mostly notable by Open Grid Forum (OGF) with Open Cloud Computing Interface (OCCI) [7], DMTF's Open Virtualization Format (OVF) [8] and etc.

There are few related research work in the area of the cooperation between NGN and cloud computing. ETSI Technical Committee (TC) GRID considers interoperability

gaps between grid/cloud technologies and the NGN architecture in an integrated environment and proposes solutions to the identified gaps [9]. Reference [10] discussed the possibility to integrate NGN/IMS and Cloud computing and offered two interconnection scenarios for combining Cloud and NGN applications in a unified architecture. Reference [11] introduced an architecture where IaaS is the development platform used for building the IMS system.

III. ARCHITECTURE OF IMS-BASED CLOUD COMPUTING

A. Architecture Overview

This paper presents an IMS-based cloud computing platform that can provide cloud services under the control of IMS. Fig. 1 illustrates the functional architecture of cloud computing based on IMS, which extends the current IMS specification with the required functionality to meet additional requirements of cloud services. IMS provides an open and standard service platform and performs the uniform service control with the subscriber service profile. The functional architecture contains main functions and reference points. The main functions associated to cloud services can be divided into two parts: cloud service functions and cloud interaction functions. The core IMS forwards the complete SIP signaling for session management and service notification of cloud services. The data flows of actual interaction between UE and cloud platform do not traverse the core IMS. This architecture supports the deployment of cloud services in a multi-provider environment.

The UE communicates with the cloud platform over multiple interfaces for different purposes, namely, over a Gm interface via the IMS core for the session management and service perception purpose, over a Ut interface for user profile configuration purpose, and over a Xd interface for actual interaction with cloud platform purpose. These interfaces are compatible with 3GPP IMS specifications.

User data involved in providing cloud services can be classified into two categories: IMS profile and cloud specific profile. IMS profile includes all information required to establish IMS sessions and access cloud services hosted in application servers. Cloud specific profile encompasses all information required to operate a cloud service, such as list

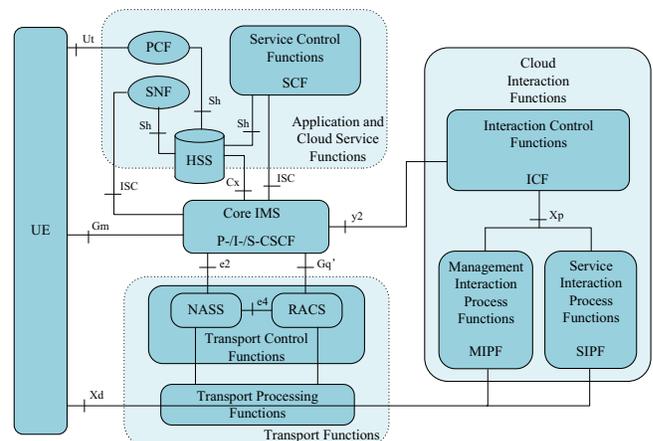


Figure 1. Functional architecture of IMS-based cloud computing

IMS common service capabilities, such as presence and group management, could be exposed to and reused by the cloud services via standardized interfaces. In the same way, some basic cloud services, such as Amazon EC2 and S3, can also be reused to create more complicated cloud services. The end-users only need one authentication for access to all authorized IMS services including various cloud services, and obtain the contracted quality of experience.

Cloud Application Servers (or ASs for short) running with SCFs, SNFs and/or PCFs manage and control the interaction between cloud clients and cloud platform. From the perspective of IMS core network, the Cloud ASs are the portal to the corresponding cloud platform. The deployment of Cloud ASs makes the actual interaction between cloud clients and cloud platform under the unified control of standardized signaling mechanisms of IMS and thus provides possibility for the use of IMS QoS and charging mechanism. Cloud Interaction Servers (or ISs for short) running with ICFs and/or IPFs.

IV. SEVERAL KEY ISSUES

A. Cloud Notification Service

The SIP events framework [12] defines general mechanisms for subscribing to, and receiving notifications of, events with SIP systems. This paper defines a new SIP event package for cloud notification service which is for subscription and notification of cloud service information. The new event package is similar to the existing event package such as user presence and call state.

After successful registration process with IMS core, UE sends a SUBSCRIBE message with cloud event package to Cloud SNF AS. In a multi-provider environment, a common Cloud Notification Application Server can be deployed. As shown in Fig. 3, the cloud clients can SUBSCRIBE to a common domain URI such as sip:domain.com or a dedicated URI such as sip:cloudservice@domain.com, and be connected to the cloud notification service provided by the common Cloud Notification AS. The Cloud SNF AS associated to a cloud platform makes use of PUBLISH message to perform the state publication of all cloud services

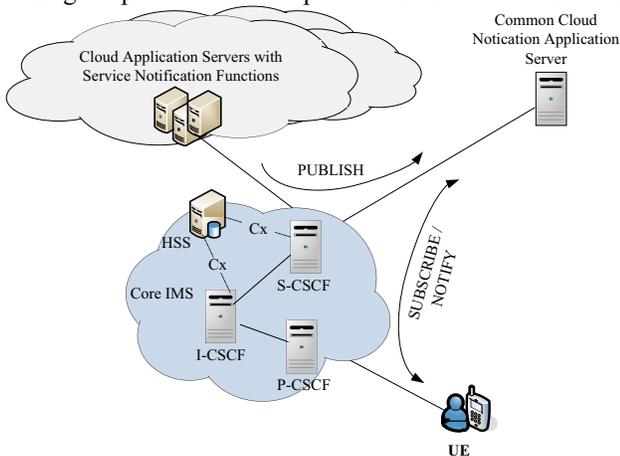


Figure 3. Overview of cloud notification service

provided by the associated cloud platform. All subscriptions to, and publications of, cloud event package are routed to the common Cloud Notification AS. The common Cloud Notification AS combines the separate cloud event partial states published by multiple Cloud SNF ASs into a whole cloud event state and acts as the notifier distributing cloud state to the interested cloud clients in the form of SIP NOTIFY messages with an XML payload. When a new cloud service provided or an existing cloud service removed, this state changes and is reported through the notification service. Through this mechanism, cloud users will learn all information of available cloud services offered by the whole domain and select the appropriate cloud service according to the state information received and their own demand, such as charging policy, the preferred service vendor.

This paper defines Cloud Information Data Format (CIDF) as a common cloud service data format and also defines a new media type “application/cidf+xml” to represent the XML MIME entity for CIDF. In cloud event package, the body of the PUBLISH/NOTIFY requests contains a cloud service information document which is formatted into “application/cidf+xml” media type.

Fig. 4 gives the non-normative diagram of the overall hierarchy of CIDF. A cloud service information document begins with the root element tag <cloudservices> which is comprised of zero, one or more <cloudservice> child elements. Each <cloudservice> element describes a single cloud service and is comprised of <service-description> and <service-state> child elements. The <service-description> element describes the static information of a cloud service such as service type (IaaS, PaaS or SaaS), service vendor, service URI (be used in order to access the cloud service), the charging policy and so on. The <service-state> element indicates the dynamic information of a cloud service such as the current user count and whether the cloud service is currently active.

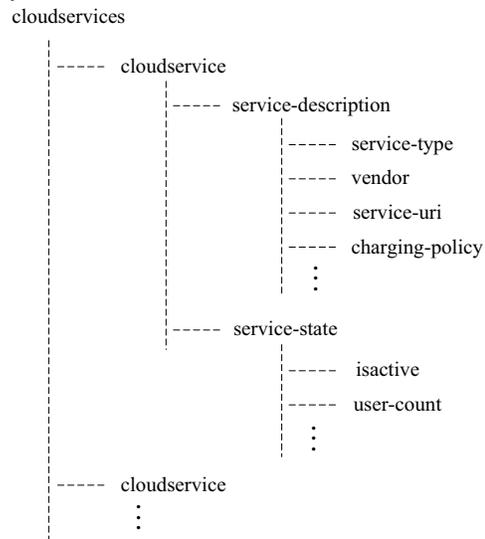


Figure 4. The non-normative diagram of the overall hierarchy of CIDF

B. Media Negotiation and QoS Control

3GPP introduced the Policy and Charging Control (PCC) architecture [13] that integrates flow based charging, QoS and policy control since release 7. Although IMS works at the application signaling level, thanks to the PCC architecture, IMS achieves negotiating the transport level parameters with the network operator infrastructure. In IMS, the P-CSCF generates the service information from the application specific media description (e.g. SDP media description). The four critical components for mapping an SDP description into a QoS authorization are the media announcements ("m="), the connection data ("c="), the attributes ("a=") and the bandwidth ("b=").

In order to establish the transport channel for cloud interaction, the media negotiation of session establishment needs to be extended to support new media types and formats. SDP is extensible easily to support new media types, formats and attributes, and has so far defined audio, video, text, application and message media types. In addition, to provide negotiable and differential QoS for packet transmission of cloud services, the signaling process of SIP session for the cloud interaction should also carry the related QoS requirements determined by the cloud interaction so that sufficient network resources can be reserved in advance. Cloud management interaction has the following transmission features: high reliability, low data rate, insensitivity to delay, interactivity, etc. Here we give an example of SDP extension for the cloud management interaction using OCCI as the standardized cloud interface protocol. In aspect of transporting OCCI protocol, SIP messages tend to be relatively small and are not fit for direct exchange of bulk data of OCCI interface. Appropriate mechanisms for traffic of OCCI protocol include the Hypertext Transfer Protocol (HTTP), the Message Session Relay Protocol (MSRP), or other media plane data transport mechanisms. The OGF OCCI group has released the OCCI HTTP Rendering document defining the RESTful OCCI API. And this document specifies three content types to represent the data being transferred: text/occi, text/plain and text/uri-list. Correspondingly, we extend two new media level attributes, "location" and "accept-types" in SDP. The location attribute indicates the path of the root of the OCCI implementation where is located the query interfaces which must be implemented by all OCCI implementations. The accept-types attribute indicates the content type of the data. In addition, the b attribute should be used to determine the maximum authorized data rate. Meanwhile, we use the Connection-Oriented Media (COMEDIA) [14] specification for setting up and maintaining reliable connections as part of the negotiation mechanism. Table 1 is an example of part of a SDP body offered by Cloud SCF AS in 200 OK response message to INVITE request message.

The different types of cloud service interactions, such as file transfer and online movies, require different network status parameters including transmission rate, delay, error rate, etc. It is seemingly impossible to formulate a unified SDP description format for all types of cloud service interactions. But from another point of view, each type of cloud service

TABLE I. AN EXAMPLE OF PART OF SDP BODY

Part of SDP Body
c=IN IP6 cloud.example.com
m=application 8080 TCP/HTTP *
b=TIAS:32
a=accept-types:text/plain text/occi text/uri-list
a=location:http://cloudas.example.com:8080/users/
a=setup:active
a=connection:new

interaction can be classified into one of media types (conversational voice/video, interactive gaming, TCP-based, etc) defined by the 3GPP PCC architecture according to the transmission requirements of resource type (guaranteed bitrate or non-guaranteed bitrate), priority, packet delay budget and packet error loss rate. And in PCC architecture, the Policy and Charging Rules Function (PCRF) maps the service information received from P-CSCF into common IP QoS parameters (e.g. the QoS Class Identifier, the data rates) for all types of media sessions, including audio/video conversation as well as typically TCP-based services and applications. Therefore, SDP only needs to be extended to carry related information for the PCRF to be able to identify the appropriate QoS class identifier, the data rates and other IP QoS parameters. The UE can retrieve necessary network parameters from cloud service information received from Cloud SNF AS and take them as the QoS parameters of the corresponding cloud service.

C. Charging for Cloud Service Usage

IMS provides the capacity of correlating charging at network and service level and supports a variety of charging modes, such as online and offline charging, time based and volume based charging, and so could meet various charging requirements of cloud services. According to the requirements of IMS charging system, Cloud SCF ASs and Cloud ICF ISs need to implement a Charging Trigger Function (CTF) and act as the charging client. The CTF is responsible for monitoring service usage and generating corresponding charging events. The charging architecture of cloud services integrated in IMS is shown in Fig. 5. The charging system triggers the charging process and sends a request to the offline or online charging functions. The Cloud SCF ASs and Cloud ICF ISs are able to distinguish

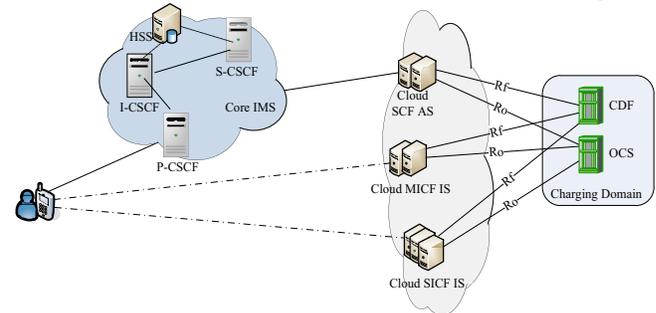


Figure 5. The charging architecture of cloud services integrated in IMS

- [5] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," RFC 3261, June 2002.
- [6] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, et al. "Above the clouds: a Berkeley view of cloud computing," Technical Report EECS-2009-28. University of California, Berkeley, 2009.
- [7] Open Cloud Computing Interface. <http://occi-wg.org/>
- [8] Open Virtualization Format. <http://www.dmtf.org/standards/ovf>
- [9] ETSI TR 102 659-2, "GRID; study of ICT grid interoperability gaps; part 2: interoperability gaps and proposed solutions," V1.2.1 (2009).
- [10] Fabricio Carvalho de Gouveia, Richard Good, Thomas Magedanz, Neco Ventura, "The use of NGN/IMS for cloud and grid services control and management," Southern African Telecommunication Networks and Applications Conference (SATNAC 2009), Mbabane, Swaziland, August 2009.
- [11] J. L. Chen, S. L. Wu, Y. T. Larosa, P. J. Yang, Y. F. Li, "IMS cloud computing architecture for high-quality multimedia applications," 2011 7th International Wireless communications and mobile computing conference (IWCMC 2011), Istanbul, Turkey, pp. 1463-1468.
- [12] A. B. Roach, "Session Initiation Protocol (SIP)-specific event notification," RFC 3265, June 2002.
- [13] 3GPP TS 23.203 v11.6.0, "Policy and charging control architecture," 2012.
- [14] D. Yon and G. Camarillo, "TCP-based media transport in the Session Description Protocol (SDP)," RFC 4145, Sep. 2005.
- [15] 3GPP TS 32.260 v11.4.0, "IP Multimedia Subsystem (IMS) charging", 2012