A Performance Based Evaluation of SIP Signalling across Converged Networks

Adetola Oredope and Antonio Liotta

***Abstract*—As wireless networks evolves towards the 3G and 4G architectures the mobile core networks provides a platform for all-IP convergence of mobile and fixed networks allowing for better integration of the Internet and cellular networks, providing better quality of service, charging mechanisms and integration of different services. Due to this convergence, there have been various evidences of interoperability limitations between the mobile and fixed networks such as protocol adaptation, end-to-end Quality of Service (QoS), charging and signalling interworking. Most of these limitations are currently been addressed by respective standards bodies and various proposals are being implemented. In this paper we investigate the signalling limitations across converged networks which are caused by the use of different Session Initiation Protocol (SIP) profiles for signalling in both mobile and fixed networks. We carried out various performance based simulations on the 3GPP proposed solution in which our results reveal critical shortcomings in the solution. We then proposed a new configuration for the system to enable the solution work more efficiently**

***Index Terms*— SIP, B2BUA, IMS, 3G, NGN**

I. INTRODUCTION

Next Generation Networks (NGNs) aims at providing a wide range of services to end-users over an access-independent platform while allowing for better Quality of Service (QoS), charging mechanisms and integration of services [1] as compared to conventional fixed or mobile networks. The NGN core network is known as the IP Multimedia Subsystem (IMS) which is the Third Generation Partnership Project (3GPP) standardised core network for the all-IP convergence of fixed and mobile networks [2].

The IMS is a service-centric platform providing a wide range of rich services such as Voice over Internet Protocol (VoIP), Push-to-Talk, Click-to-Dial, Online gaming or an integration of these services to provide a personalised end-user experience. It is also built on various internet protocols such as the Internet Protocol version 6 (IPv6) [3] for addressing, IP Security (IPSec) [4] for network security, Session Initiation Protocol (SIP) [5] for signalling, session management and session control, Common Open Policy Service (COPS) [6] for Quality of Service policing and Diameter [7], for authentication, authorization and accounting.

The core functionality of the IMS is built on the Session Initiation Protocol (SIP), the Internet Engineering Task Force (IETF) standardised protocol for the creation, management and termination of multimedia sessions on the Internet [4].

SIP is a highly matured protocol but notwithstanding, there are still certain limitations with the SIP implementation such as scalability [7], redundancy [8], QoS [9], security [10], Network Address Translation (NAT) / Firewall transversal [11] and node mobility [12] to mention a few. Most of these issues are addressed using SIP extensions which allow the “Require” and “Supported” fields of the SIP message headers to easily add new features to the protocol and still allow for backward compatibility with existing SIP systems.

Over fixed networks, the use of SIP extensions is not compulsory but to allow for the full functionality of SIP in mobile networks, especially in cellular environments, the 3GPP recommends the use of certain SIP extensions such as the Integration of Resource Management and SIP [8], the SIP UPDATE method [9] and the Reliability of Provisional Responses in SIP [10]. This is known as the 3GPP profile of SIP which tends to vary from the conventional SIP that is used in fixed networks. This introduces a limitation in converged networks in which SIP messages needs to transverse across fixed and mobile networks

In this paper, we evaluate the performance of the 3GPP proposed solution for the above limitation which is based on the implementation of a Back-to-Back User Agent (B2BUA) [11] for interworking the 3GPP profile of SIP (3GPP SIP) used in mobile networks and the conventional profile of SIP (non-3GPP SIP) used in fixed networks. This was achieved by carrying out various performance-based tests under different scenarios to determine the impact of the B2BUA on end-to-end services between users across the converged network. Our results reveal certain drawbacks in the system such as the introduction of various delays and the loss of relevant signalling packets in the system. Based on these results, we then proposed an improvement in the system that will allow for the B2BUA to work more efficiently.

Section II of this paper gives a brief description about SIP

Manuscript received March 23, 2007. This work is being undertaken in the context of IST-034115 PHOSPHORUS (Lambada User Controlled Infrastructure for European Research – http://www.ist-phosphorus.eu

Adetola Oredope is a Research Assistant with Department of Electronic Systems Engineering, University of Essex, Wivenhoe Park, CO4 3SQ, United Kingdom. (phone: +44 (0) 1206 87-4247; fax: +44 (0) 1206 87-2900; e-mail: aoredo@essex.ac.uk).

Dr. Antonio Liotta is a Senior Lecturer with the Department of Electronics Systems Engineering, University of Essex, UK (e-mail: aliotta@essex.ac.uk).

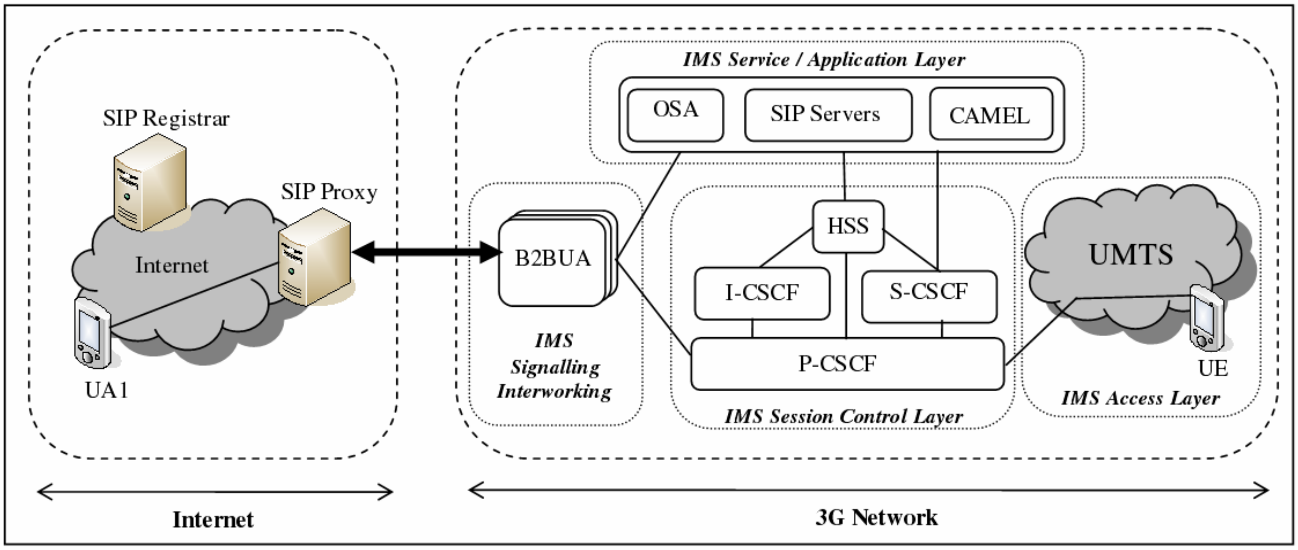

Fig. 1. SIP Signalling across a Converged Network

with a focus on the use of SIP over converged networks. Section III describes our IMS test bed, performance tests and results. Section IV then provides our proposed configurations. Finally, our conclusions and future recommendations are provided in Section IV.

II. BACKGROUND TECHNOLOGIES

A. *Session Initiation Protocol (SIP)*

Multimedia communications over the internet are built on the Session Initiation Protocol (SIP) [5], proposed by the Internet Engineering Task Force (IETF) as the protocol for handling call setup, routing, authentication and other feature messages to endpoints within an IP domain [12]. SIP is also used to deliver the session descriptions using the Session Description Protocol (SDP) [13] to allow the participants in a session agree on a set of compatible media.

A SIP network consists of a User agent (UA) that can create new SIP requests (User Agent Client (UAC)) and can also respond to SIP requests (User Agent Server (UAS)), that is, a UA is a logical entity that can act as both a UAC and a UAS [5]. A SIP UA can be implemented in either software or hardware. In Fig. 1, the SIP UA is a Voice IP (VoIP) Phone connected directly to the Internet.

The SIP Proxy server (Fig. 1) is used in forwarding SIP requests received from UAs or other proxies to other locations (proxies, UAs or external networks). It is also used to authenticate and authorize users which are already registered their current locations with the SIP Registrar [12]. The base SIP specifications defines six SIP requests which are also know as methods and they are INVITE, ACK, OPTIONS, BYE, CANCEL and REGISTER [5].

SIP is based on HTTP and it maintains a standard format which is made up of the start line, header fields and the message body. In order to address some of the limitations of SIP some UAs normally apply the use of SIP extensions [12], which is a negotiation process takes place between UAs to find out which extensions are supported by both UAs and which extensions are actually needed by session [14]. This is done by using the "Require" and "Supported" header fields. The UA adds the extensions needed in the "Require" header of a message and also lists the extensions it supports in the "Supported" header message. This is examined by the other UA and a reply is sent with its own required and supported extensions.

The use of SIP extensions on the Internet is optional but the all SIP elements must support the core SIP methods. The implementation of SIP extensions depends on the equipment vendor, network operator or the end-user. The use of SIP extension has allowed new services to be created such as Instant Messaging [15], Location Services [16] and Presence Services [17].

B. *SIP in Converged Networks*

Converged networks merges both fixed and mobile networks to provide a multi-access platform, blending both fixed and mobile capabilities and providing a seamless and personalised service to end users. At present, 3G networks have been able to deploy this service-centric architecture based on the IP Multimedia Subsystem (IMS) core network which uses SIP as its core enabler for session management and control. SIP allows clients to INVITE other clients into a session; it then negotiates control information, terminal capabilities and media channels to be used in the session, IMS then allows the clients to be connected via IP.

The main functions of IMS are to allow for authentication and authorization of both fixed and mobile clients, support roaming, charging, and Quality of Service (QoS), providing access and network domain security [1]. IMS also facilitates the deployments of Internet based applications such as Voice over IP (VoIP), Instant Messaging, Push-to-talk, Video Conferencing, Presence and Online Gaming in 3G networks.

As shown in Fig. 1, the IMS core consists of the Call Session Control Function (CSCF), a SIP Server responsible for the registration of the end points and the routing of the SIP signalling messages. There are three types of CSCF which are the Proxy CSCF (P-CSCF), interrogating CSCF (I-CSCF) and the Serving CSCF (S-CSCF). Other IMS functions also include the Home Subscriber Server (HSS) database which maintains the unique end-user service profile, gateways and application. Unlike using SIP on the Internet, where SIP extensions are optional, the 3GPP mandates the use of core SIP, SIP extensions and other protocols to achieve the desired level of session control and management in the IMS network [1]. This can be described as a profile because it does not differ from the usage of SIP over the internet but the 3GPP has mandated a number of SIP extensions and options in both the IMS network nodes and the IMS terminals [1]. This is known as the 3GPP SIP profile and it is described in 3GPP TS 24.299 [18]. The 3GPP SIP profile allows the IMS to provide better QoS, integration of services and billing to the end-users by specifying the minimum requirements needed to set up a session and this can be a major problem if the IMS needs to interwork with SIP networks external to the IMS such as the Internet as explained in the next section.

C. Interworking SIP between Fixed and Mobile Networks

One of the major advantages of converged networks is that mobile and fixed users can easily communicate and use end-to-end services over a single platform using SIP as explained in sections 2.1 and 2.3. But due to the different profiles of SIP used in both mobile and fixed networks it is evident that there will be an interoperability limitation. Based on this SIP interworking limitation, the 3GPP specifies in 3GPP TR 29.962 [11] the use of a Back-Back User Agent (B2BUA).

A B2BUA is a logical entity that can act as both a UAC and a UAS to modify, generate and respond to SIP messages [5]. Unlike a proxy server, it maintains dialog state and must participate in all requests sent on the dialogs it has established [14].

The 3GPP specifies that a B2BUA is permanently inserted at connections between the IMS and a given external network [11] handling all SIP signalling, including session attempts, subscriptions, instant messaging and also signalling where the flows may forward without B2BUA intervention. Although the B2BUA is inserted between the IMS and the external network, it only become active when it receives a SIP “420” (Bad Extension) response with the "Unsupported" header field including the "preconditions" tag [8] from a non-3GPP UA.

The B2BUA also stores the SDP offer in initial INVITE requests for all calls until receiving a provisional or final response from the Non-3GPP UA [11]. Once all the requirements in both networks are fulfilled, the B2BUA then processes the messages and in most time, the SIP messages are re-written to allow the networks understand the messages. In some cases, the SDP is also examined and the RTP body is also converted to a format that is understood at the network it is designated for. The 3GPP also recommends that the B2BUA must be updated regularly in order for it to be able to support all extensions and required media parameters.

Although the B2BUA solution provides a platform for SIP interworking, we believe that there are major drawbacks in the solution. A major draw back is that the B2BUA breaks end-to-end paradigm of SIP between nodes during which some additional forms of delay are introduced into the system. These delays may cause time sensitive applications such as VoIP to break. Also, the amount of delay introduced into the system is made up of various parameters such as the examination of SIP messages, re-writing of SIP messages and re-initiation of sessions after an handover to mention a few. These delays may seem negligible individually but on a commercial scale when millions of SIP messages are involved this will be noticeable and also cause a degradation of the overall system. This motivated us to carry out the various experiments in Section 3 in order to investigate the real impact of the B2BUA on the overall system.

III. EXPERIMENTAL TESTS AND ANALYSIS

A. Test Bed

The test bed is implemented on the Linux 2.6 kernel with a user environment of Fedora Core 5- Linux. The kernel supports various protocols such as IPv6 and IPSec out-of-the-box which makes the deployment of next generation services possible. The architecture of our test bed varies for different experimental scenarios with the B2BUA interconnecting the IMS and the SIP network.

To allow for proper performance analysis and tests, we used the Hewlett-Packard (HP) SIPp [19] application to develop our B2BUA components. SIPp is a free and open source SIP test tool and Traffic generator, to inject SIP traffic into the network at varying conditions. It also allows for the use of XML files to customize SIP call flows and scenarios supporting both TCP and UDP protocols. SIPp can also send RTP traffic and can also provide statistics on the running tests while other parameters such as call rate, call duration and round trip delay can be easily adjusted. SIPp also support IPv6, Transport Layer Security (TLS) and SIP authentication.

Based on SIPp functionality we developed our B2BUA components, UAS and UAC in XML and were integrated with the SIPp application to run the tests.

In order to develop our B2BUA, we implemented a UAS and a UAC to communicate directly via SIPp. The XML based scenario for the UAS listened and processed SIP traffic generated by the UAS and provides appropriate responses to the UAC just as described Section 2.3. The UAS generated random requests that were varied to emulate different experimental scenarios as explained in the tests later on. The B2BUA was implemented using an Intel Pentium III (Coppermine) processor running at 1GHz speed and 512MB random access memory.

The performance measurements of the CPU and Memory of the B2BUA while processing the SIP messages were collected using Top [20], a Linux command that provides an ongoing look at processor activity in real time, listing the most CPU-intensive tasks on the system. It can sort out tasks based on CPU usage, memory usage and runtime. Top can also be used to output its contents to files when used in the batch mode.

B. Test 1: Platform Validation

1) Aim

These set of tests aim to validate our IMS platform especially the signalling interworking ability of the B2BUA to rewrite, manage and process SIP messages.

2) Setup

The B2BUA implementation was started using the XML based scenarios of the UAS and UAC. The source and destination addresses were configured using the local host address in order to allow the real-time measurement of performance without the introduction external effects such as network parameters in order to prevent message re-transmissions and timeouts. A total of 80 calls were made at the rate of 10 calls per second.

3) Result

The total call time was 8.04 seconds with all calls successfully placed and all messages appropriately exchanged. This shows that our test bed is fully functional and there are proper SIP message interactions between the UAS and UAC components of the B2BUA.

C. Test 2: Performance Tests

1) Aim

The aim of these tests is to study and investigate the amount of resources consumed by the B2BUA in rewriting and processing SIP messages. This enables us determine the performance of the system under different load conditions.

2) Setup

The B2BUA implementation was started using the UAC and UAS XML configuration and the localhost as source and destination addresses as explained in section 3.2. All calls were also configured for duration of 10 seconds. The system was then subject to an incremental call rate based on 100 steps between 0 calls per second and 1000 calls per second with each step lasting for about 20 seconds. Linux Top was also used in batch mode to collect the CPU and Memory information of the system during the tests.

3) Results

Based on the tests above, the system behaved normally between 10 calls per second (cps) and 600 cps after which the performance behaviour changed as shown in Fig. 2.

From Fig. 2 the CPU processed 100 cps with as little as 1.0% of the total CPU usage with corresponding Memory usage of 1.6% as also shown in Fig. 3. At this point all calls were fully processed and the appropriate re-writing of SIP messages was achieved. It is also important to note that the CPU utilization increased by 5% for every 100 steps increment of calls per second until about 500cps as shown in Fig 2. During this period, the memory utilization also increases at the rate of 1.5% for every 100 steps increment in calls per second.

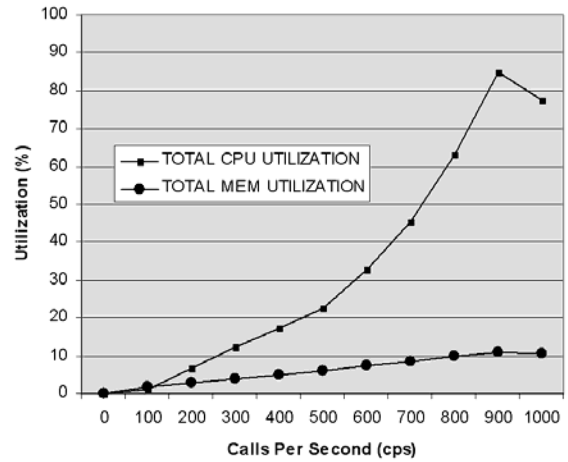

Fig. 2. Total CPU and Memory Utilization

After this period (0-500cps), the CPU utilization tends to increase faster as the rate of about 10% for every 100cps increment due to fact that more messages were generated. We also noticed that due to the increase in the number of messages, the response times of the B2BUA also started to increase and the UAC component of B2BUA started generating retransmissions. This increased the total amount of messages in the system which in turn rapidly increased the CPU utilization from 32% at 600cps to about 84% at 900 cps. At this point the CPU reached its saturation point in which it stopped processing all messages and started to discard the messages. This in turn reduced CPU utilization to about 77% at 1000cps but still discarding all messages as shown in Fig. 2. The CPU performance began to degrade at this point.

All through the period of 600cps to 1000cps, the memory utilization continued to increase at the rate of 1.5% for every 100cps increment as shown in Fig. 2 and was not affected by the increase in messages. The memory utilization also all started to decrease between the 900cps and 1000cps region.

We believe that this behaviour exhibited by the memory is due to the fact the memory size is fixed and it can only hold a fixed amount of messages for the CPU to process at a particular time. As the messages increase, more messages arrive at the CPU than it can process whilst leading to a high CPU utilisation (Region 600cps-900cps). As the CPU holds other incoming messages in a queue in the memory, when the memory limit is reached additional messages are discarded.

Comparing the CPU and Memory utilization before 900 cps, the CPU utilization increased in a logarithmic scale while the Memory utilization increased with a linear scale. This helps us to understand that the re-writing of SIP messages in B2BUA is a lot more CPU intensive as compared to memory utilization.

The results can also help in the generation of analytical models that simulate real-life scenarios.

D. *Test 3: Fault Tolerance Tests*

1) *Aim*

This section aims to investigate and study the CPU behavior under highly stressed conditions. This will enable us understand the fault tolerance properties of the B2BUA and its process of recovery from a fully saturated conditions.

2) *Setup*

In order to investigate the fault tolerance properties of the B2BUA, we stressed the system by configuring a total of 900 calls per second with each of 240 seconds call duration. The whole system was then studied for about 4 minutes.

3) *Results*

The results collected during the experiment are shown in the graph in Fig. 3. The B2BUA CPU utilization at the beginning of the test is about 11% for 900cps which then starts to increase gradually until it reaches roughly twice its incremental rate of 1% for every 3 seconds to about 43% CPU utilization. We believe that at this moment the CPU utilization was increased rapidly due to the increase in messages. This allows the CPU quickly process any outstanding messages and allows it to return to normal operation. The normal operation continued until 1:17 minutes into the experiment in which the CPU utilization increased to about 88%. From this moment, the CPU began to discard messages rapidly in order to accommodate new messages. This degradation of the system continued for until about 2:35 minutes into the experiment, after which all messages had been discarded and the CPU utilization became 0% i.e. no messages were being processed. From this point the B2BUA tried recovering but the UAS and UAC were out of synchronization and could not process corresponding messages accordingly. The whole system finally collapsed after 3:33 minutes.

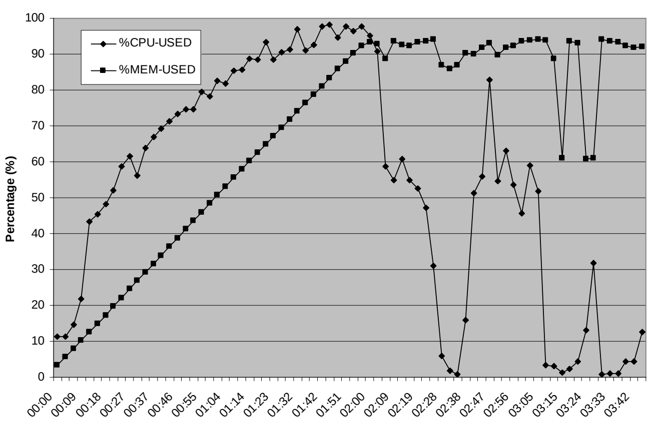

Fig. 3. Fault Tolerance Results

Based on this experiment, it is evident that B2BUA has major problems recovering from saturated points especially when its UAC and UAS components are out of synchronization. The process of discarding messages at saturation point is a major draw back on the system.

IV. Proposed Architecture

The tests in Section III shows that the major draw back on the test bed is the CPU performance as shown in Fig. 2 and 3. This is caused by the processing power demanding task of re-writing the SIP messages as they transverse the B2BUA.

Furthermore, as the number of SIP messaging increases, the systems further degrades leading to continual loss of packets. In order to reduce the amount of time used in processing the calls and continually re-writing the SIP messages we suggest an additional component in the B2BUA known as the Session Mapping Database (SMD).

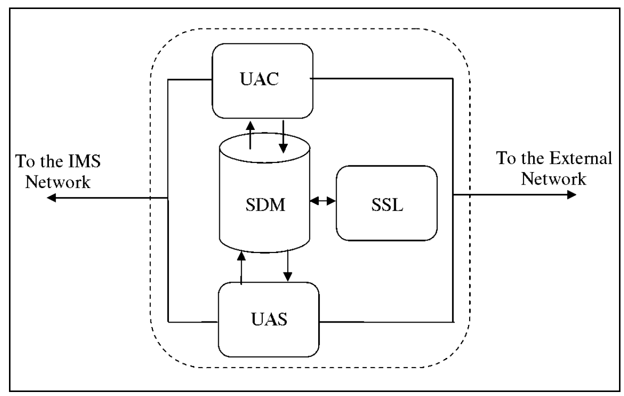

Fig. 4. Proposed approach for B2BUA Implementation

The SMD is a distributed and redundant database that is used to store the mappings between 3GPP SIP profiles and non-3GPP SIP profiles that successfully transverse the B2BUA. The SMD database is initially populated with the corresponding mappings between the 3GPP UA and the non-3GPP UA. The main advantage of doing this is that the amount of CPU usage is significantly reduced and can be used for other processes. The mappings are applied automatically to subsequent sessions with exactly the same session parameters, preventing all SIP messages between the IMS and external networks to be re-written continually. The SMD database is then replicated within the IMS system to allow a smooth recovery from any failures encountered by the B2BUA. This approach is shown in figure 4.

Once a message is received in the B2BUA, it is passed to the SDM to examine the SIP and SDP message patterns. If the pattern exists in the SDM, then the required mapping is applied and the messages are passed to the required parties. On the other hand, if the pattern is a new one, it is passed to the SSL to re-write the SIP message to the appropriate one, 3GPP or non-3GPP. The pattern and corresponding mappings are then passed back to the SDM to be stored and used for subsequent SIP messages following the same patterns.

The SDM can be implemented based on any distributed

ISBN:978-988-98671-2-6

database and can also be replicated with other Operation and Management databases in the IMS.

V. Conclusion

In this paper we have investigated the effects of the B2BUA in SIP interworking between the fixed and mobile networks. Our results show that the tasks assigned to B2BUA are CPU intensive which in turn causes SIP messages to be discarded during times of high load making the solution an inefficient one. We also noticed that most of the delay introduced into the system is based on the processing time of the CPU re-writing the SIP messages and as the number of SIP messages that transverse the B2BUA increases, the performance of the systems degrades.

We also proposed an architecture that we believe will improve the system significantly However, the interworking of the fixed and mobile networks is inevitable and it provides a platform for rich and personalised services combined with mobility advantages.

These are the first stages of our implementation of an effective content delivery platform in converged networks. We aim to integrate the Peer-to-Peer (P2P) paradigm of the Internet into the Client-Server architecture of converged networks. This allows for the advantages of P2P such as scalability, redundancy and availability to be achieved over an effective content delivery platform in 3G Networks.

References

[1] Camarillo, G., et al.: The 3G IP multimedia subsystem (IMS) : merging the internet and the cellular worlds. J. Wiley, Chichester, West Sussex ; Hoboken, NJ (2006)

[2] (3GPP), T.G.P.P. IP Multimedia Subsystem (IMS) Specification Release 5. http://www.3gpp.org/specs/releases-contents.htm#3GRelease5 (September 2003)

[3] IETF RFC 2460: Internet Protocol, Version 6 (IPv6) Specification.

[4] IETF RFC 2401: Security Architecture for the Internet Protocol.

[5] IETF RFC 3261: SIP: Session Initiation Protocol.

[6] IETF RFC 2748: The COPS (Common Open Policy Service) Protocol.

[7] IETF RFC 3588: Diameter Base Protocol.

[8] G. Camarillo, E., et al. IETF RFC 3312: Integration of Resource Management and Session Initiation Protocol (SIP). (October 2002)

[9] Rosenberg, J. IETF RFC 3311: The Session Initiation Protocol (SIP) UPDATE Method. (September 2002)

[10] Rosenberg, J., et al. IETF RFC 3262: Reliability of Provisional Responses in the Session Initiation Protocol (SIP). RFC 3262 - Internet Engineering Task Force (June 2002)

[11] 3GPP-TR29.962 3GPP TR 29.962: Signalling interworking between the 3GPP profile of the Session Initiation Protocol (SIP) and non-3GPP SIP usage (Release 6). (October 2005)

[12] Camarillo, G.: SIP demystified. McGraw-Hill, New York (2002)

[13] Handley, M., et al. IETF RFC 2327: SDP: Session Description Protocol. (April 1998)

[14] Sinnreich, H., et al.: Internet communications using SIP : delivering VolP and multimedia services with session initiation protocol. Wiley, New York (2001)

[15] Campbell, B., et al. IETF RFC 3428: Session Initiation Protocol (SIP) Extension for Instant Messaging. RFC 3428 - Internet Engineering Task Force (December 2002)

[16] Rosenberg, J., et al. IETF RFC 3263: Session Initiation Protocol (SIP): Locating SIP Servers. RFC 3263 - Internet Engineering Task Force (June 2002)

[17] Roach, A.B. IETF RFC 3265: Session Initiation Protocol (SIP)-Specific Event Notification RFC 3265 - Internet Engineering Task Force (June 2002)

[18] 3GPP-TS24.229 3GPP TS 24.229: IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3 (Release 7). (June 2006)

[19] HP SIPp - Test tool/Traffic Genrator.

[20] Top : Linux / Unix Command.