Research Article

Effective Performance Metrics for Multimedia Mission-critical Communication Systems

Ashraf Ali[1],* and Andrew Ware[2]

[1]The Hashemite University, Jordan
ashraf@hu.edu.jo
[2]University of South Wales, UK
andrew.ware@southwales.ac.uk
*Correspondence: ashraf@hu.edu.jo

Abstract: Mission-critical Communication Systems that are adaptable for use with the latest generation of multimedia services are crucial for system users. To determine the set of requirements that need to be hardcoded into such systems, a clear distinction between mission-critical and non-mission-critical systems is required. Moreover, the users of services provided by such systems are very different to those of current mobile commercial communication systems. These differences give rise to a set of challenges that need addressing to facilitate migration from existing systems to those now being proposed. One such challenge relates to the performance of the IP Multimedia Subsystem (IMS) registration process. This is a crucial consideration for mission-critical systems, particularly in large-scale systems where thousands or even millions of users may seek to access the system in disaster scenarios. This paper presents an evaluation of IMS and Session Initiation Protocol (SIP) performance metrics and Key Performance Indicators (KPIs). Moreover, it articulates a proposed study that will seek to address some of the challenges identified.

Keywords: *IMS; IP Multimedia Subsystems; Key Performance Indicators; Long Term Evolution; LTE; Performance Metrics; Session Initiation Protocol; SIP*

1. Introduction

'Mission Critical' is defined by TETRA and The Critical Communication Association (TCCA) as a function whose failure would lead to catastrophic consequences that would place public order or public security at risk. A system that provides such critical functionality must have suitable inbuilt functionality, interoperability, security, and the wherewithal to maintain its availability. Mission-critical users are those with responsibility for the welfare, health, security and safety of the public. Law enforcement forces, firefighters, emergency and medical services, rescue services, military forces, utility staff members, and transport services members, are all example of mission-critical users. The concept of a Mission-critical Communication Systems (MCCS) refers to the hardware, software and communication facilities, that allow mission-critical users to communicate with each other and liaise with command centres securely and dependably for the sake of providing mission-critical services, wherever and whenever the services require special communication solutions.

Many governmental and non-governmental organisations (NGOs) play critical roles in the operation of mission-critical systems. The services provided by such organisation are of a type that makes them less tolerant to execution or operational errors. Such services are referred to as mission-critical services. The criticality of such services implies that they have a set of special requirements that distinguish them from other services. They should be available anytime, anywhere, within the

service operational scope. Moreover, they need to be able to function at the extremity of their capabilities regardless of the operational circumstances or running conditions.

The most important protocol performance issues that control the operation of end-to-end systems and have implications on its overall performance are presented in this paper. There are mainly two types of protocols: control plane protocols and data plane protocols. The Session Initiation Protocol (SIP) is one of the control plane protocols and was fully standardised by the Internet Engineering Task Force (IETF) in RFC 2543 for the first version of SIP 1.0 and RFC 3261 for the second version of SIP 2.0 [1]. The SIP is a communication protocol that operates over IP and is used for the signalling of real-time multimedia services such as voice, video and non-real-time services (for example, text messages and presence notifications). The text-based protocol mainly defines the signalling order between end-users for call initiation and termination, in addition to instantly modifying the call setup as needed. It is also used for registering the users before the call is initiated.

Although the SIP message headers and signalling details are not presented in this paper, performance issues and the challenge of enhancing SIP services performance are briefly discussed.

2. SIP Performance Issues

The protocol related performance metrics need to be identified to determine how the SIP is utilising the system resources and to help ensure optimal usage. Moreover, the architectural design challenges need to be targeted to enhance the SIP performance. A variety of tests that measure the processing time of SIP messages, memory allocation, thread performance, and call-setup time was presented [2]. The results show that the performance of the proxy server is affected by varying SIP related parameters, thus affecting the number of calls per second that the proxy server can handle. The network topology, such Wireless Sensor Networks (WSNs) affect the SIP performance requirements [3]. The results also show that the performance of SIP related architectures, such as the IP Multimedia Subsystems (IMS) presented later, are more affected than a simple Proxy/Registrar Server due to the heavy dependence of SIP signalling and messages structure. Furthermore, it is important to note that the performance of SIP signalling is significantly affected by the delay at different stages of registration, call initiation, and call termination processes. The performance of SIP signalling also affects the QoS of the offered service. Hence, the need to define the metrics that identify the performance measure for the SIP is crucial for evaluation and performance comparison purposes. The IETF proposed the criteria for the end-to-end SIP performance measures in RFC 6076 [4].

Understanding SIP performance is essential for determining the overall multimedia system communication performance. IMS, for example, is composed of many entities that communicate with each other using SIP signalling. In other research, the impact of the scheduler settings on the overall performance of multithreaded servers was investigated [5]. The work showed that the operating system and the hardware architecture of the machines that host the SIP server have a significant impact on the SIP performance. Service scheduling for multithreaded systems, for example, is controlled via the process scheduler that ensures that the multi-process and threads within the SIP server are serviced fairly and that all available resources are utilised appropriately. The Completely Fair Scheduler (CPS), for example, is used in Linux based systems to schedule the threads that need to be serviced by the CPU, based on certain fairness criteria. The study relied on using specific CPU performance-related parameters to alter the default operation of the CPS to achieve better scheduling performance and therefore better SIP server performance.

Due to the lack of a SIP benchmarking standard to define the baseline performance of SIP signalling, RFC 6076 was developed to articulate the performance metrics for SIP in VoIP applications to provide Key Performance Indicators and Service Level Agreement (SLA) indicators for best network resource utilisation and best end-user Quality of Experience (QoE). As shown in figure 1, the primary metrics defined in RFC 6076 are the Registration Request Delay (RRD), Session Request Delay (SRD), and Session Disconnect Delay (SDD). RRD is the time needed for the user to finish the

registration process successfully. SRD is the time required to get a reply from the server-side regarding the requested call setup from the user side; it is counted for both successful and unsuccessful call requests. If the call requested were successfully set then the call setup time would simply be the SSD plus the acknowledgement sending time. The SDD is the time difference between sending the BYE message from the user side and the time of receiving 200OK confirmation from the server.
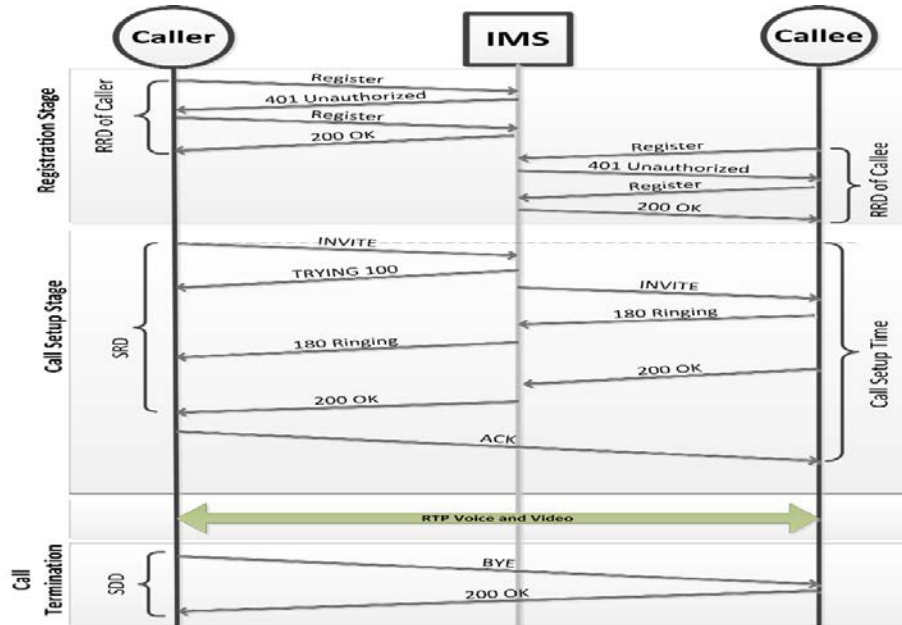

Figure 1. SIP signalling flow and performance metrics.

In this paper, the call setup delay is used to measure the system performance due to its importance in real-time multimedia services in general and in mission-critical communications specifically. The QoE for SIP-based systems is affected significantly by the call setup value. The mean time of call setup values can reach up to 800 ms [6]). However, for LTE-based mission-critical systems (designed to work as a replica for traditional dedicated mission-critical systems such as TETRA), the average call setup time needs to be within 500 ms delay. Figure 2 shows the average access time of different technologies. The call setup time is usually a multiple of the access time due to the enforced handshaking process.
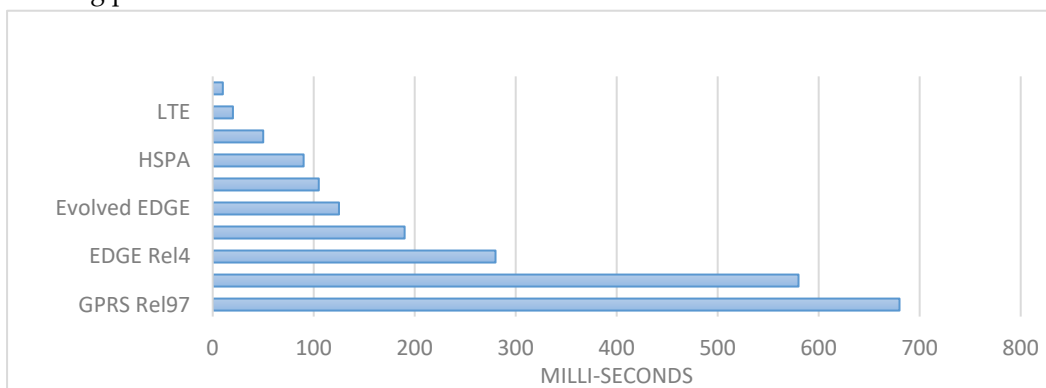

Figure 2. Access Time Evolution for Different Technologies

3. IP Multimedia Subsystem (IMS)

Legacy mobile communication networks and the next generation systems can provide end-users with new sets of applications and services. There is a need for a system that sits on top of the access technology domain to provide the required signalling for multimedia services. IP Multimedia Subsystems (IMS) form the core part of Next Generation Networks (NGNs). They play an important role in helping the service operators merge the multimedia services in cellular networks and to

provide end-users with key features and services within Quality of Service (QoS) levels set by operators. These multimedia services such as messaging, instant voice, video conferencing, group management, and push services, all rely on IMS to control the signalling and the state of the service before initiation, during the service and after service termination. IMS can be integrated in non-multimedia based systems as well [7].

IMS was designed and standardised by 3GPP to provide multimedia services over mobile communication technologies beyond GSM [8]. The IMS is used for delivering the IP multimedia services between users and service providers; it gains its importance as an architectural framework for multimedia communications, it can also be integrated in other frameworks for robot messaging and logging [9]. The SIP is the principal signalling protocol used within an IMS to create, modify and destroy multimedia sessions. As defined by the standard, an IMS functions as an interface between the service/application layer and the transport layer, enabling the service providers and operator to control the user QoS based on its subscription profile. Moreover, it works as a hub point for the entire SIP signalling that needs to take place before, during, and after the call. To ensure overall integrity, the different control functions are connected through a set of interfaces.

Users' subscription-related information is stored in the Home Subscriber Server (HSS) that performs authentication and authorisation functions for users. The HSS is also responsible for updating user registration status records. The Call Session Control Functions (CSCF) are responsible for handling the SIP signalling messages and packets in the IMS. The Proxy CSCF (P-CSCF) is the entry point into the IMS system, and all SIP messages flow through the P-CSCF. The P-CSCF may also apply security or compression algorithms to the received traffic in addition to providing quality of service control and bandwidth management. The interrogating CSCF (I-CSCF) is one of the main elements of the IMS systems. It is used during the registration process when the UE does not know which Serving CSCF (S-CSCF) should receive the request. The I-CSCF interrogates the HSS to obtain the address of the appropriate S-CSCF that should process the request.

The S-CSCF performs session control functions that interface with the HSS to check and download the user profile information. It assigns the Application Server (AS) for the user for further services and enforces the operator policy control. The SIP AS has a SIP interface with the S-CSCF and is used to host specific IMS services. After the registration process is completed and the S-CSCF is allocated to the UE, the I-CSCF is no longer used for any further communication. All future communication happens between the UE, the P-CSCF, and the S-CSCF.

In the next sections, the performance of IMS, which is critical in affecting the overall system performance due to its hierarchal position in the stack and core functional role, is discussed.

3.1. IMS Performance Issues

After migrating from the circuit-switched Second Generation Mobile Networks (2G) toward the packet-switched domain of fourth-generation (4G) communication networks and beyond, the need for supporting multimedia services in all IP network infrastructure is essential to ensure the convergence of data, multimedia services, and mobile networks technologies. IP Multimedia Subsystems (IMS) developed by the Third Generation Partnership Project are a key part of Next Generation Networks (NGN) that are responsible for providing and controlling the multimedia services in the packet-switched domains. As defined by the standard, the IMS operates as an interface between the service/application layer and the transport layer that enables the service providers and operator to control the user QoS based on its subscription profile. Moreover, it works as a hub point for the entire SIP signalling that needs to take place before, during, and after the call. For this purpose, there are different functions that are connected by interfaces to ensure operational integrity.

In the protocol stack, the IMS resides between the application layer and the transport layer (see figure 3). Therefore, the performance of the IMS affect mission-critical applications that have strict requirements due to their special operating nature and associated tasks. End-end delay of the access technology components along with the IMS entities affect the end-to-end signalling, and QoE offered to the end-user. Having a generic system that has IMS core network serving both legacy mobile users

and mission-critical system users will introduce overload, especially at the IMS side. Statistical analysis of the IMS queues helps in measuring the system performance [10].
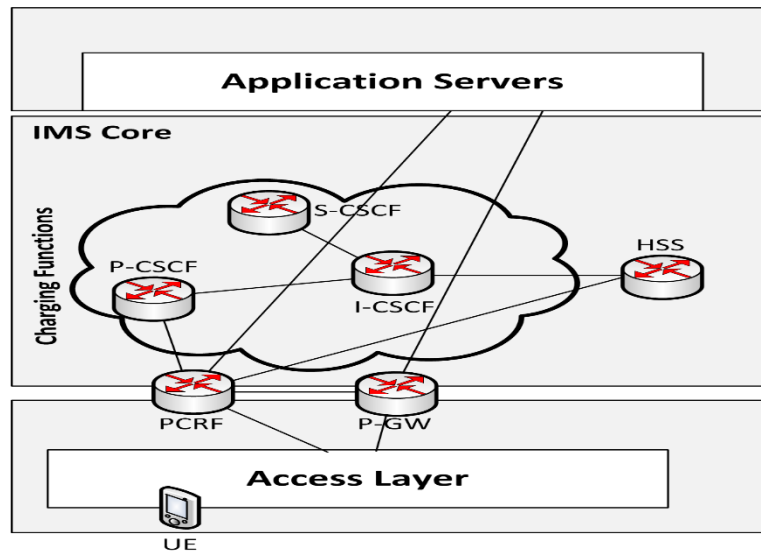

Figure 3. IMS Charging Functions

The performance of the IMS is affected by the individual performance of each entity inside it. The P-CSCF is the first entity that interacts with the User Element (UE) and forwards the SIP messages to other control functions in the IMS. It is also used to apply security or compression algorithms to the received traffic in addition to monitoring quality of service and bandwidth management. The I-CSCF is a SIP server that is assigned by the HSS to the user when the I-CSCF requires it. The S-CSCF has interfaces with the HSS to perform session control, download the user profile information, assign the Application Server (AS) for further services, and to enforce the operator policy control. The SIP AS has a SIP interface with the S-CSCF, and is used to host specific IMS services. Some gateways function as interfaces with other domains.

3.2. IMS Registration Performance

In the IMS registration, the user requests authorisation to gain access to the IMS services. Figure 4 shows the registration steps in IMS. First, the UE sends a SIP Register request to the P-CSCF, which forwards it to the I-CSCF. Then the I-CSCF sends a User Authentication Request (UAR) to the Home Subscriber Station (HSS), using the Diameter protocol, to check the user profile for authorisation process and to determine the S-CSCF address allocated to the user. Then, the HSS replies with the User Authentication Answer (UAA) to the I-CSCF and authorises the user. After retrieving the S-CSCF details from the HSS, the I-CSCF forwards the SIP Registration request to the S-CSCF, which in turn sends a Multimedia Authentication Request (MAR) to download user authentication data to the HSS. The HSS then replies with the Multimedia Authentication Answer (MAA) to the S-CSCF, which in turn responds to the user using a SIP 401 Unauthorised response that embed an authentication challenge within it for the UE. The UE then generates another SIP Register request following the same steps as described earlier. This time the authentication process finishes and the S-CSCF sends a Server Assignment Request (SAR) to the HSS, using the Diameter protocol, which replies with a Server Assignment Answer (SAA) using the same protocol. Finally, the S-CSCF sends a SIP 200OK message to the UE to complete the registration process.

As mentioned before, the Registration Request Delay (RRD) is the time needed for the registration process to complete. In this case, that is the time difference between the first SIP Register message sent by the UE and the time the SIP 200 OK message is received. Figure 5 shows the transition diagram of the IMS Registration process.
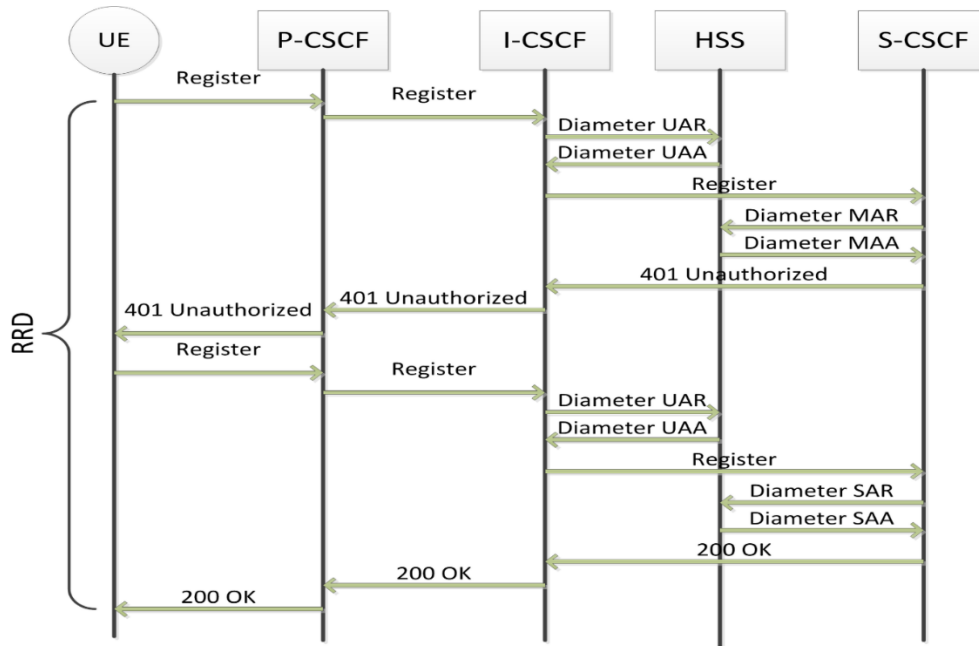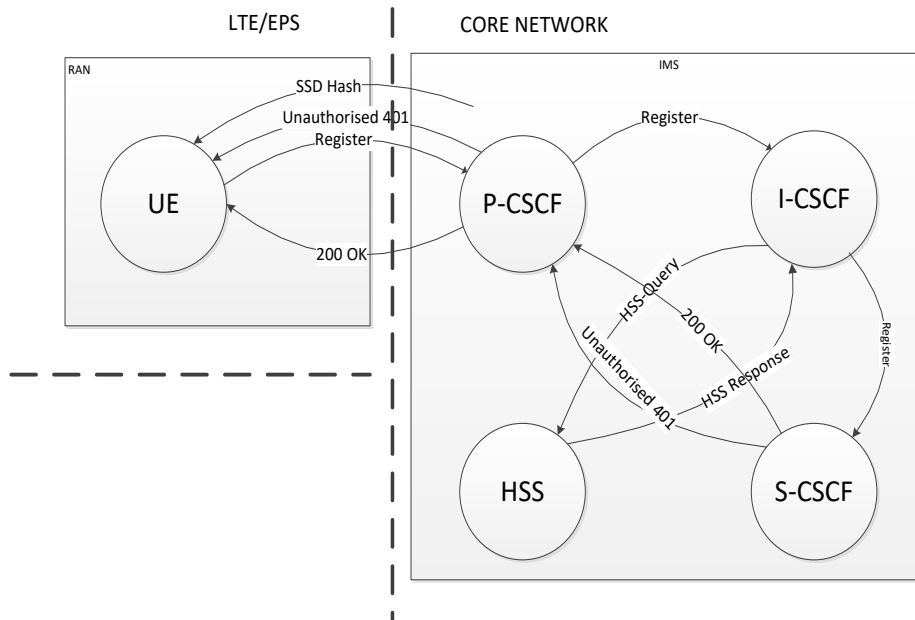
Figure 4. IMS Registration Process

Figure 5. Transition Diagram of IMS Registration Process

During the registration process, the end-user needs to be subscribed to the IMS to be authorised to make calls and initiate other services instantly during the call. The registration process is required before starting calls (see figure 6) and reregistration is needed for the SIP device to re-authenticate itself after receiving a notification from the SIP server. This is required to avoid compromising the SIP device.

The delay in the registration process is one of the SIP performance metrics and criteria that was defined by IETF in RFC 6076 [4], due to the previous lack of a SIP benchmarking criteria to define the baseline performance of SIP signalling. RFC 6076 defines the performance metrics for the SIP in VoIP applications in order to provide key performance indicators and Service Level Agreement (SLA) indicators for best network resources utilisation and best end-user Quality of Experience (QoE). The main metrics defined in RFC 6076 are the Registration Request Delay (RRD), Ineffective Registration Attempts (IRA), Session Request Delay (SRD), and Session Disconnect Delay (SDD). RRD is the time needed for the user to finish the registration process successfully and plays a major role in the overall QoS.
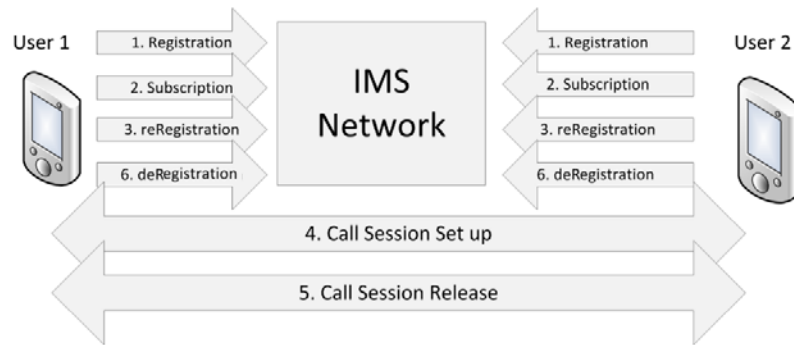
Figure 6. User Registration with IMS

In addition to the delay in the registration process, the rate at which registration messages are sent to the server is of great importance. This is especially true in mission-critical operations where the number of users or devices that request access to the SIP/IMS server may be high enough to affect or even deteriorates the performance of the service. Different real-world scenarios for registering users/devices over LTE have been presented by other researchers. For example, in one scenario, an Airbus A380 carrying a thousand passengers was simulated after landing at Heathrow airport, with each passenger initiating IMS Registrations over the LTE at a rate of ten registrations per second. In the second scenario, a regional power outage was restored at a big city causing a flurry of IMS registration requests to the network, a hundred registrations per second was expected in this scenario [11]. In both scenarios, it was important to handle all the requests successfully and finish the registration process within acceptable time limits regardless of the received traffic. This successful registration is especially important in mission-critical system that by their nature, should be always available, resilient and reliable.

PoC is one of the mission-critical services supported by IMS. It is designed to operate as a replica of Push To Talk (PTT) service in conventional dedicated mission-critical communication systems such as TErristrial Trunked RAdio (TETRA) networks and P25. Therefore, the investigation of the IMS added delay in the registration process is of great importance for mission-critical services which are less delay tolerant, especially if operating over a generic communication system that is not dedicated to mission-critical communications.

Some studies that analysed IMS registration over WiMax and 3G interworking architectures show that the access technology along with IMS design affect the registration signalling delay [12]. In one study, a performance analysis of IMS signaling, including registration signalling, in multimedia networks was carried out. The study showed that the maximum signalling delay is affected by both the number of User Elements (UEs) and the available bandwidth available in the core network. However, the study was limited to a maximum of 400 users [13]. Other researchers have proposed a general IMS registration protocol for wireless networks in order to reduce the registration delay by combining both the registration processes with the access technology and IMS network [14]. A reregistration procedure modification was proposed in order to reduce the transmission delay in the IMS [15]. Reregistration is needed either in order to refresh the registration state of the user or to update the user registration profile due to change in user capabilities. Moreover, when the impact of SIP signalling load over IMS was investigated, it was found that the SIP signalling load, including registration, has a significant impact on the Key Performance Indicators (KPI) for IMS [16].

This paper investigates the delay in the registration process introduced by the IMS for increasing numbers of UEs. Moreover, it also articulates the effect of increasing the load over the KPI and related QoE.

4. SIP Key Performance Indicators

SIP key performance indicators are important to measure end-to-end and core network performance, Software-Defined Networking (SDN) and Network Function Virtualization (NFV) rely

on the KPIs for better controlling the network [17]. A set of SIP KPIs defined by 3GPP include the following. The Registration Request Delay (RRD), which is the difference between the time of final successful registration arrival and the time of sending the registration request. The number of Ineffective Registration Attempts (IRA) is a metric that indicates the ratio of unsuccessful registration attempts to the total number of registration requests sent. The Session Request Delay (SRD) is the time difference between the reception of status indicative response and the time of sending the SIP invite message. The Session Disconnect Delay (SDD) is the time difference between sending the SIP Bye message and the reception of confirmation or timeout message. The Session Duration Time (SDT) is the time interval between the receptions of the 200OK SIP message following the INVITE Request and the receipt of the BYE timeout message at the originating UE. The Session Establishment Ratio (SER) defines the ability to terminate the User-Agent or Proxy Server during the session establishment. The Session Establishment Effectiveness Ratio (SEER) is the ratio of SIP Invite Requests that end with a 200OK response from the terminating side and the number of SIP INVITE Requests that end with a response type other than a 200OK. The Ineffective Session attempts (ISAs) is the ratio of failed session setup requests and the total number of session requests. Finally, the Session Completion Ratio (SCR) is the ratio of successful completed sessions and the total number of sessions. This paper considers both the RRD and IRA as KPI metrics that reflect the performance of the SIP Registration process.

5. IMS Key Performance Indicators

The criteria for carrying out IMS performance evaluation using test-beds has been defined by the European Telecommunications Standards Institute [18]. Moreover, 3GPP in their technical standard [19] have defined a set of KPIs for IMS that can be classified in to three categories:

A) Accessibility KPIs: which are a set of metrics that reflect the accessibility of the IMS for the users. They are:

1) Initial Registration Success Rate (IRSR) of S-CSCF: reflects the accessibility performance provided by the IMS and measures the percentage of attempted user initial registrations (UR.AttInitReg) that were successful (UR.SuccInitReg). The percentage is calculated as follows:

$$SR = \frac{\sum_{s-cscf} UR.SuccInitReg}{\sum_{s-cscf} UR.AttInitReg} * 100\% \tag{1}$$

2) Mean Session Setup Time: measures the meantime for session setup starting from when the SIP Invite message is sent to the P-CSCF until the reception of the 200OK.

3) Session Establishment Success Rate: is calculated by the I-CSCF using two values:- the ratio of successful session initiation requests to the total number of session initiation attempts; the ratio of successful session termination attempts compared to the total session termination attempts.

$$SR\_Orig = \frac{\sum_{Type} SC.succSessionOrig.type}{SC.AttSessionOrig} \tag{2}$$

$$SR\_Term = \frac{\sum_{Type} SC.succSessionTerm.type}{SC.AttSessionTerm} \tag{3}$$

The type notation in the nominator stands for SIP 180 and SIP 200OK message types.

4) Third-Party Registration Success Rate: this KPI is obtained by dividing the number of successful third party registration procedures by the attempted third party registration procedures.

$$TPRSR = \frac{\sum_{s-cscf} UR.Succ3rdPartyReg}{\sum_{s-cscf} UR.Att3rdPartyReg} * 100\% \tag{4}$$

5) Reregistration Success Rate of S-CSCF: the ratio of successful reregistrations of S-CSCF to the total number of attempted reregistrations.

$$PRSR = \frac{\sum_{s-cscf} UR.SuccReReg}{\sum_{s-cscf} UR.AttReReg} * 100\% \tag{5}$$

6) Mean Session Setup Time Originated from IMS (MSSTOI): the mean setup time of successful originated calls from the IMS.

B) Retainability KPI: it has one value, the Call Drop Rate of IMS Sessions, which is calculated as the ratio of dropped sessions to the number of successful session established.

$$SEDR = \frac{SC.DroppedSession}{\sum_{Type} SC.succSession.type}$$ (6)

C) Utilisation KPI: it has only one value, the Mean Session Utilization (MSU), which is calculated as the percentage of the mean number of simultaneous online answered sessions compared to the maximum number of sessions provided by the IMS network.

$$MSU = \frac{SC.NbrSimulAnsSessionMean}{Capacity} * 100\%$$ (7)

The QoE, which measures the quality of the service as perceived by end-users, is affected by the KPI values of both the SIP and IMS. In order to improve the QoE, the corresponding KPI values need to be continuously monitored and improved. The KPI measure, if combined together, will help in determining the failure points in the IMS network and be indicative of how to improve things so as to avoid the bottleneck and single point failure in case of traffic overload scenarios.

The KPI may also be used as a Load Detection Function (LDF). The LDF may be used to support load balancing for the S-CSCF and, therefore, reduce the service request latency.

6. Voice Applications

There are different types of applications and data formats used in mission-critical communication system. These include voice, video, and push messages, IP-based voice and video applications over SIP is widely used and deployed in many systems [20]. Some of these applications are termed "killer applications" due to their bandwidth requirements, especially at the core and access point side, which are considered bottlenecks in the overall system. Another application, Push To Talk (PTT) enables the caller to push a button to initiate the sending of a voice burst to the callee side. This particular application uses half-duplex channels to save bandwidth and minimise the need for extra bandwidth. There is a need for addressing users scalability concerns that may emerge due to two-way or full-duplex communication methods which will be a challenge for future broadband mobile communication systems if PTT over Cellular (PoC) are adopted in the future. However the adoption of PoC introduced additional delays especially when applied over LTE which experience more delays that need to be analysed [21].

Table 1 VoIP total Data Rate for Different Codecs **[22]**

Codec	G.711	G.729	G.723.1[A]	G.722.1[B]	G.722.1[C]	EVRC[E]	EVRC[F]	EVRC[G]	AMR[H]	iLBC[I]
Sample Time (ms)	10	10	30	20	20	20	20	20	20	20
Frame Size [J]	640	80	158	480	640	172	80	16	95	303
Packets Per Second	100	100	33	50	50	50	50	50	50	50
IP,RTP,UDP headers	320	320	320	320	320	320	320	320	320	320
Ethernet Header[K]	176	176	176	176	176	176	176	176	176	176
Total Size	1136	576	654	976	1136	668	576	512	591	799
Ethernet CS data rate[L]	113.6	57.6	21.80	48.8	56.80	33.4	28.8	25.6	29.5	39.9
IP CS data rate[L]	96	40	15.9	40	48	24.6	20	16.8	20.7	31.1
A. 5.3 Kbps **B**. 24Kbps **C**. 32 Kbps **D**. 48 Kbps **E**. Rate 1 **F**. Rate ½ **G**. Rate 1/8 **H**. 4.75 **I**. 15.2 Kbps **J**. All sizes measures in bits **K**. Including the CRC and 802.1Q header **L**. Calculated in Kbps										

Voice over IP (VoIP) is used for the delivery of voice bursts over IP networks. Due to the emergence of All-IP-Networks, VoIP has gained significant importance as the best voice streaming solution for voice calls. Voice Codec is used at the subscriber side to convert the analogue voice waves to digital pulses and vice versa. There are different codec types based on the selected sampling rate, data rate, and implemented compression algorithm. In order to determine the bandwidth requirements of VoIP connections, there is a need to identify common VoIP Codecs and their

associated characteristics (table 1 shows the Bandwidth Requirements of different Codecs [22]. VoIP services are considered bandwidth-killing applications for mobile broadband Access Technologies that provide end-users with services similar to those of DSL providers in the traditional wired networks. However, the wireless access mechanism introduces additional challenges and constraints on the data rate as well as delay requirements. Therefore, a bandwidth requirements study for VoIP was needed to ensure that the mission-critical service that use VoIP operate as expected and to determine how best to increase the scalability of the networks within acceptable QoS levels. A comparison between the real the theoretical codec values is important indication of the quality of the call and network performance [23].

7. Multimedia Services in Mission-Critical Systems Challenges

Based on the earlier discussion, there are two different deployment options for the mission-critical systems, either to have a dedicated system or a commercial system. Unfortunately, there are drawbacks for both options and many trade-offs that control the design and deployment process. What is important to note is that at the nascency of dedicated mission-critical systems, voice and simple text messages were the only services used by PPDR for their critical missions. Simply during the early stages of the evolution of mobile communication networks, operators only thought of providing voice communication between users. Hence, when dedicated mission-critical communications were deployed, the standardisation bodies designed the whole system only for voice and low data rate services. However, due to the exponential increase of mission-critical data used to support different services and tasks, the need for broadband data along with the traditional basic voice service has become vital for the mission-critical communication systems.

Due to the increased demands for more bandwidth, the need for an access technology that is capable of providing broadband connectivity and scalability, along with broadband network support in the core network, is of crucial importance for any mission-critical communication system.

7.1. Access Technology Challenges

As discussed earlier, there is an increasing demand for more bandwidth for mission-critical services. This implies that systems not only need to satisfy the general requirements of mission-critical services, but also a more scalable system to serve more users. Moreover, the users' broadband services should be running as it should be during crises and worst-case scenarios.

Increasing the bandwidth of the system is not a simple task in the physical layer of any communication system especially as there are multiple factors and trade-offs that complicate the evolution process of any communications system. From a business perspective, mobile network operators need to utilise every Hertz of the invested licensed frequency band to help ensure maximum profit. Technically speaking, the frequency spectrum should be utilised so to achieve the highest possible bandwidth to support the broadband services. Hence, the need to deploy methods for the best frequency exploitations such as: adaptive modulations, Multiple Input Multiple Output (MIMO) Antennas, Orthogonal Frequency Division Multiple Access (OFDMA), Code Division Multiple Access (CDMA), Frequency Division Duplex (FDD), Time Division Duplex (TDD) and many other technologies in both physical and MAC layers.

There are trade-offs between the frequency, transmission power, coverage area, and cell size. In addition, there is a trade-off between the frequency and maximum system throughput based on the Bit Error Rate value. Finally, the system throughput controls the downlink and uplink capacity of the resource blocks, which limits the maximum number of simultaneous users and the scalability of the whole system.

Current operational 4G mobile communications networks are widely deployed using the Long Term Evolution (LTE) standard articulated by 3GPP. While the LTE standard can be part of a public safety system, it is important to note that previously existing trade-offs still cause issues within the access technology abstract. However, the LTE or other technologies dedicated technologies can provide reasonable solutions in the physical and MAC layers to maximise the capacity of the system.

7.2. System Resources Sharing Challenge

Another challenge is using commercial mobile communication networks as a public safety communication system. As mentioned before, there are two options - either a dedicated system or commercial system for the mission-critical services. Clearly, it is a significant challenge to enable a general-purpose public, commercial communication system to work as a replica of a dedicated public safety communication system with all the mission-critical services and strict requirements that have to be satisfied. On the one hand, using the dedicated system will enable all the communication system resources to be exclusively available to the PPDR users. On the other, for the commercial, public safety communication system, the resources will be shared between the PPDR and public, commercial mobile users. Therefore, for the public communication system, the need to have a distinction between the mission-critical services and other non-mission-critical services in terms of service priority handling and resource allocation. This ensures that the system will have enough resources and thus be available for the PPDR users whenever and wherever needed.

7.3. System Architecture Design Challenge

Mission-critical communication systems are composed of different entities and interfaces to ensure end-to-end connectivity and overall system integrity. End-end users may be connected to different network domains and access technologies, hence the need for interfaces and gateways that provide transparent connectivity between the different domains. In addition, there is a need for standardised interfaces that connect devices regardless of manufacturer. Finally, there is a need for data plane and control plane protocols that ensure the connectivity and data delivery between end-users.

Figure 7 shows the general layered architecture model. Regardless of the details of any mission-critical communication system, all systems share the same hierarchal architectural structure. End-users are the ones holding the mobile device that is running the applications that interact with the required services through the available access technology. The access technology used to reach the core network can be either a single predetermined access technology or multiple access technologies. Clearly, having more than one supported access technology to access the core network will support the reliability requirement of requested mission-critical service.

Regardless of the access technology used, the need to have a common platform working between the access technology and the mission-critical application servers is a necessity for two reasons. Firstly, the platform needs to work as intermediary between the control signalling and data streaming signalling domains. Secondly, the platform is needed to provide seamless connectivity for the end-users along with the application servers regardless of the access technology being used. Moreover, it should be noted that there is no clear distinction between the data domain and control domain signalling protocols.
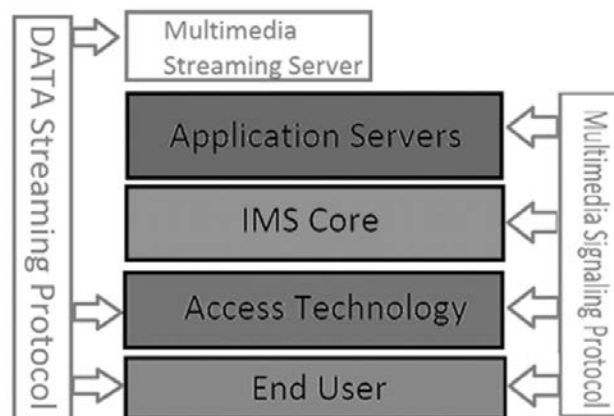

Figure 7. Layered Abstract Model

7.4. End-to-End QoS Challenge

Having different abstracts, in a hierarchal manner for the architecture of the system as described previously, will introduce not only integration challenges but also cross-layer service performance optimisation challenges. Such structure is important to improve the quality of service for the whole end-to-end mission-critical communication system and to satisfy the mission-critical requirements mentioned previously.

7.5. System Performance Modelling and Validation Challenge

Modelling a system and simulating its behaviour, while changing a set of dynamic parameters, is useful in providing an understanding of its overall functionality and responses. Verification and validation are essential parts of the readability and availability testing process.

The verification process of a mission-critical system starts with a model design (or framework) analysis. The system design needs to be checked against all specifications before the system is implemented and built. For a newly proposed system, the technical reports and the standardisation bodies in the field are responsible for setting the guidelines and set of specifications need to be followed by system engineers. This helps ensure a more realistic and accurate model. The lack of finalised standards imposes many challenges in the verification and validation steps.

8. Proposed Solutions

The challenges outlined in the previous section highlight the need for a further study to focus on the cross-layer communications between different system abstracts. The result of this study should help ensure that calls, such as mission-critical emergency calls, with higher priority class, pass through the different entities of the system toward the end-user while at the same time satisfying all mission-critical service requirements. In addition, there is a need to develop a pre-emptive scheduling algorithm capable of deciding which type of calls need to be dropped in order to free up allocated resources at times when system capacity is being exceeded. To facilitate the development of such a pre-emptive scheduling algorithm, a thorough evaluation of overall system capacity with different scenarios that combine a number of users along with the priority class of the service will need to be undertaken.

The proposed study will need to develop a framework for the mission-critical communication system. Moreover, the different layers and interfaces for both data plane and control plane need to be demonstrated. Furthermore, the layered architecture model of the system and the details for each layer need to be investigated and clarified.

A cross-layer optimisation and integration of the signalling used in both control and data domains need to be proposed as part of a general framework. Finally, in order to test the proposed framework and the overall proposed system architecture in addition to validating the proposed cross layer optimisation, the whole system needs to be modeled using simulation tools. The Quality of Service performance measures need to be compared with the performance metrics of a test-bed for the mission-critical system that will be used as a benchmark.

One of the main objectives for this study is to propose a cross layer optimisation solutions between both the interfaces and protocols running in the mission-critical system to help facilitate the enhancement of the overall performance and its associated metrics.

9. Conclusions

This paper articulates some of the many challenges that need to be addressed with respect to current commercial mobile communication systems. The main goal is migrating to a generic system that is better in terms of performance than current dedicated mission-critical communication systems and, at the same time, more scalable to serve increasing number of commercial mobile users. The performance of SIP and IMS, as a core part of the generic proposed system, has significant implication on the system's overall performance. The registration process, as an example, is considered a good

way to measure the system response time as most of the system entities are involved during the registration process. Similar to the registration, there are a set of KPIs for both the IMS and SIP that can be considered as a reference points to measure the system performance at different layers and abstracts of the end-to-end system. The challenges and requirements that need to be addressed satisfied in any proposed alternative have also been articulated.

References

[1] Rosenberg Jonathan, Schulzrinne Henning, Camarillo Gonzalo, Johnston Alan, Peterson Jon et al., "SIP: session initiation protocol", Request for Comments: 3261: pp.1-269 (2002), DOI: 10.17487/RFC3261.

[2] Mauricio Cortes, J. Robert Ensor and Jairo O. Esteban , "On SIP performance", Bell Labs Technical Journal, vol. 9, no. 3, pp. 155-172, 2004, DOI: 10.1002/bltj.20048.

[3] Zhang Liec Lee, Bi-Lynn Ong, Amiza Amir and Wei Siang Hoh, "A survey of session initiation protocol in Wireless Mesh Network", in Proceedings of the 2017 IEEE 15th Student Conference on Research and Development (SCOReD), Putrajaya, Malaysia, 2017, pp. 286-290, DOI: 10.1109/SCORED.2017.8305377.

[4] Morton Al and Malas Daryl, "Basic Telephony SIP End-to-End Performance Metrics", Request for Comments: 6076, ISSN: 2070-1721, Internet Engineering Task Force (IETF), 2014.

[5] Ramesh Krishnamurthy and George N. Rouskas, "On the impact of scheduler settings on the performance of multi-threaded SIP servers", in Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 6175-6180, DOI: 10.1109/ICC.2015.7249307.

[6] ITU-T, "Signalling requirements for IP-QoS In telecommunication standardisation sector of ITU , (vol. series Q: switching and signalling," ITU-T, 2004.

[7] Jian-Wei Li and Yu-Hung Huang, "Integrate Non-IP Multimedia Subsystem Devices in IP Multimedia Subsystem Services", in Proceedings of the 2019 8th International Conference on Innovation, Communication and Engineering (ICICE), Zhengzhou, China, 2019, pp. 114-116, DOI: 10.1109/ICICE49024.2019.9117462.

[8] 3GPP, "Service Requirements for the Internet Protocol (IP) Multimedia Core Network Subsystem (IMS)", In Technical Specification Group Services and System Aspects, Third Generation Partnership Project, 2005.

[9] S Thejashwini, M Sunil Kumar and Sini Anna Alex, "IMS Based Session Initiation Protocol in Robot Framework for Telephony Services", in Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 1218-1223, DOI: 10.1109/ICIRCA.2018.8597315.

[10] Mauro Mario Di and Liotta Antonio, "Statistical Assessment of IP Multimedia Subsystem in a Softwarized Environment: A Queueing Networks Approach", IEEE Transactions on Network and Service Management, vol. 16, no. 4, pp. 1493-1506, December 2019, DOI: 10.1109/TNSM.2019.2943776.

[11] James Ankin, Joseph Zeto and Kathy O'Neil, Validating VoLTE: A Definitive Guide to Successful Deployments, 2nd ed. California, USA: CreateSpace Independent Publishing Platform, 2014, ISBN-13: 978-1497311008, ISBN-10: 1497311004.

[12] Arslan Munir and Ann Gordon-Ross, "SIP-Based IMS Signaling Analysis for WiMax-3G Interworking Architectures", IEEE Transactions on Mobile Computing, vol. 9, no. 5, pp. 733-750, May 2010, DOI: 10.1109/TMC.2010.16.

[13] Nader F. Mir, Sarhan M. Musa, Heng Gao and Chaitra Shivakumar, "Performance Analysis of IMS Signaling in Multimedia Networks", Information Engineering (IE), ISSN Online: 2326-5868, ISSN Print: 2326-585X, vol. 1, no. 1, PP.1-7, December 2012.

[14] Daniel Díaz-Sánchez, Davide Proserpio, Andrés Marín-López and Florina Weik Almenárez-Mendoza, , "A General IMS Registration Protocol for Wireless Networks Interworking", in Proceedings of the Wireless and Mobile Networking Conference, Berlin, Heidelberg, Germany 2009.

[15] Reza Farahbakhsh, Mohammad Varposhti and Naser Movahhedinia, "Transmission Delay Reduction in IMS by Re-registration Procedure Modification", in Proceedings of the 2008 The Second International Conference on Next Generation Mobile Applications, Services, and Technologies, Cardiff, 2008, pp. 142-146, DOI: 10.1109/NGMAST.2008.34.

[16] Jiri Hosek, Lubos Nagy, Vit Novotni, Pavel Masek, and Dominik Kovak, "Performance Analysis: Impact of Signalling Load over IMS Core on KPIs", in Proceedings of the Recent Advances in Circuits, Systems, Telecommunications and Control, ISBN: 978-960-474-341-4 137, Paris, France 2013.

[17] Ahmadreza Montazerolghaem, Mohammad Moghaddam, Yaghmaee Hossein and Alberto Leon-Garcia, "OpenSIP: Toward Software-Defined SIP Networking", in Proceedings of the IEEE Transactions on

Network and Service Management, vol. 15, no. 1, pp. 184-199, March 2018, DOI: 10.1109/TNSM.2017.2741258.

[18] ETSI, "IMS Network Testing (INT), IMS/NGN Performance Benchmark", Part 1: Core Concepts, TS 186 008-1 V1.2.1, European Telecommunication Standards Institute, 2012.

[19] 3GPP, "Performance Management (PM); Performance measurements; IP Multimedia Subsystems (IMS) (Release 11)", In Technical Specification Group Services and System Aspects, vol. v11.4.0, Third Generation Partnership Project 3GPP, September 2012.

[20] Rajendrasing Nalawade Namrata, Shikha Nema and Srinivas Yalampati, "Efficient IP-based voice & video communication through session initiation protocol (SIP)", in in Proceedings of the 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1-5, Coimbatore, India, 2017, DOI: 10.1109/I2C2.2017.8321862.

[21] Chengsui Lu "Delay Analysis of Push to Talk over Cellular (PoC) Service Solutions for Public Safety Communications Over LTE Networks", Master Thesis, University Politicnica De Barcelona, Barcelona, Available: http://hdl.handle.net/2099.1/16088 2012.

[22] Ashraf Ali A, Spyridon Vassilaras, Konstantinos Ntagkounakis, "A Comparative Study of Bandwidth Requirements of VoIP Codecs over WiMAX Access Networks", in Proceedings of the 2009 Third International Conference on Next Generation Mobile Applications, Services and Technologies, Cardiff, Wales, UK, 2009, pp. 197-203, DOI: 10.1109/NGMAST.2009.47.

[23] Martín Ortega Ortega, Gustavo Chafla Altamirano, Ciro Larco Barros and Mara Falconí Abad, "Comparison between the real and theoretical values of the technical parameters of the VoIP codecs", in Proceedings of the 2019 IEEE Colombian Conference on Communications and Computing (COLCOM), Barranquilla, Colombia, 2019, pp. 1-6, DOI: 10.1109/ColComCon.2019.8809181.