Performance Optimization of IP Multimedia Subsystem

Mlindi Mashologu

DISSERTATION.COM

Boca Raton

Performance Optimization of IP Multimedia Subsystem

Dissertation.com
Boca Raton, Florida
USA • 2010

ISBN-10: 1-59942-337-5
ISBN-13: 978-1-59942-337-1

Abstract

Performance Optimization of IP Multimedia Subsystem (IMS) : Mlindi Mashologu (Under the Supervision of Professor Rex Van Olst)

The IP Multimedia Subsystem (IMS) is the basic network architecture for Next Generation Networks (NGN) which is intended to bridge the divide between the traditional circuit switched and packet switched networks, thereby providing a single network capable of providing all service offerings. IMS is based on the IP infrastructure and it enables the convergence of data, speech and video on the same network platform. The IMS forms the basis of Fixed Mobile Convergence (FMC) where fixed-line operators are striving to provide mobile access and mobile operators are trying to provide fixed access and this is done to provide both services to a customer in a single device.

The IMS is based on Session Initiation Protocol (SIP) which is a text based protocol. The IMS will generally create additional signaling traffic in the IP based networks, so there is a need to take necessary precautions to minimize the signaling overload. My research is based on how the performance of the IMS can be improved by optimization of SIP as well as IMS elements.

The analysis and characterization of the signaling traffic generated by IMS has been performed and how the signaling traffic can be reduced by the compression of SIP using the Burrows Wheeler Transform (BWT). The queuing models of the IMS have been formulated and the mathematical approach has been used to find the impact of implementing the Hyper-Threading technology on the IMS Elements.

ACKNOWLEDGEMENTS

I would like to thank GOD ALMIGHTY for giving me the strength courage , wisdom and guidance to finish this Investigative Research.

I would like to thank my supervisor for providing me with the guidance in this research. Thank you for encouraging me when I thought that the task at hand was impossible for me to achieve.

I would also like to thank all the lectures of the School of Electrical and Information Engineering, University of Witwatersrand for all the courses that you have presented. Without these courses I would not be able to tackle this research.

Lastly, I would like to give the hugs and kisses to my wife who has been so supportive through out my journey as well as the thanks to all my friends and family for their encouragement.

Table of Contents

Acronyms

2G	:	Second Generation Networks i.e. GSM
3G	:	Third Generation Networks i.e. WCDMA
3GPP	:	3rd Generation Partnership Project
3GPP2	:	3rd Generation Partnership Project 2
AAA	:	Authentication, Authorization and Accounting
ADSL	:	Asymmetrical Digital Subscriber Line
ARI	:	Arithmetic Encoding
ATM	:	Asynchronous Transfer Mode
AS	:	Application Servers
AVP	:	Attribute Value Pairs
BGCF	:	Breakout Gateway Control Function
BWT	:	Burrows Wheeler Transform
CAMEL	:	Customized Application for Mobile Network Enhanced Logic
CDMA	:	Code Division Multiple Access
CDR	:	Call Data Records
CSCF	:	Call Session Control Function
DDoS	:	Distributed Denial of Service
DMC	:	Dynamic Markov Compression
DoS	:	Denial of Service
DSP	:	Digital Signal Processing
EDGE	:	Enhanced Data Rates for GSM Evolution
EPC	:	Evolved Packet Core
EPS	:	Evolved Packet System
FMC	:	Fixed-Mobile Convergence
GGSN	:	Gateway GPRS Support Node
GPRS	:	General Packet Radio Service
GSM	:	Global System for Mobile communications
HA	:	High Availability
HSS	:	Home Subscriber Server
HTTP	:	HyperText Transfer Protocol
I-CSCF	:	Interrogating Call Control Function
ICS	:	IMS Centralized Services
IETF	:	Internet Engineering Task Force
IM-SSF	:	IP Multimedia Service Switching Function
IM	:	Instant Messaging
IMS	:	IP Multimedia Subsystem
ISP	:	Internet Service Provider
ISR	:	Interrupt Service Routine
ISUP	:	ISDN User Part
IP/MPLS	:	Internet Protocol Multi-Protocol Label Switching
IP	:	Internet Protocol Call Session Control Function
IPTV	:	Internet Protocol Television
LTE	:	Long Term Evolution
JPEG	:	Joint Photographic Experts Group
MBLB	:	Message Based Load Balancing

MGW	:	Media Gateway
MPEG	:	Moving Picture Experts Group
MSC	:	Mobile Services Switching Centre
MSS	:	Mobile Softswitch / MSC Server
MRFC	:	Multimedia Resource Function Controller
MRFP	:	Multimedia Resource Function Processor
MTF	:	Move-To-Front
MTU	:	Maximum Transmission Unit
NAS	:	Network Access Server
NASS	:	Network Attachment Subsystem
NGN	:	Next Generation Networks
OFDMA	:	Orthogonal Frequency Division Multiple Access
OMA	:	Open Mobile Alliance
OPEX	:	Operational Expenditure
OS	:	Operating System
OSA	:	Open Service Access
P-CSCF	:	Proxy Call Session Control Function
PTT	:	Push-To-Talk
PSTN	:	Public Switched Telephone Networks
RACS	:	Resource and Admission Control Subsystem
RADIUS	:	Remote Authentication Dial-In User Service
RLE	:	Run Length Encoding
RTP	:	Real Time Transmission Protocol
S-CSCF	:	Serving Call Session Control Function
SAE	:	System Architecture Evolution
SBC	:	Session Border Controller
SCTP	:	Stream Control Transmission Protocol
SDP	:	Session Description Protocol
SMTP	:	Simple Mail Transfer Protocol
SIP	:	Session Initiation Protocol
SLA	:	Service Level Agreement
SLF	:	Subscriber Location Function
TISPAN	:	Telecommunications and Internet converged Services and Protocols for Advanced Networking
TCP	:	Transmission Control Protocol
TDM	:	Time Division Multiplexer
UA	:	User Agent
UAC	:	User Agent Client
UAS	:	User Agent Server
UDP	:	User Datagram Protocol
UE	:	User Equipment
UMTS	:	Universal Mobile Telecommunication System
QoS	:	Quality of Service
VCC	:	Voice Call Continuity
VoBB	:	Voice over Broadband
VoD	:	Video on Demand
VoIP	:	Voice over Internet Protocol
WCDMA	:	Wideband Code Division Multiple Access
W-LAN	:	Wireless Local Area Network
Wi-Fi	:	Wireless Fidelity
WiMAX	:	World wide Interoperability for Microwave Access

List of Figures

List of Tables

Chapter 01
Introduction

1. Introduction

The emergence of the Next Generation Networks (NGN) has changed the way the telecommunication is being seen. The IP network on the core of the NGN enables the convergence of completely different network architectures in both wired and wireless worlds and this convergence is referred to as Fixed-Mobile Convergence (FMC). The IP Multimedia Subsystem (IMS), as the core network and the Session Initiation Protocol (SIP) form the basis of triple play services (Voice, Data and Video) over the same network architecture. The IMS Multimedia Telephony service is seen as the next step in telecommunications and the services which would finally harmonize telephony with data communications [4].

The term NGN is used to describe the network that is integrated, based on open network architecture which has the ability to provide voice, data and multimedia services. The IMS is used for the Next Generation communication networks which provides interoperability i.e. multi-vendor, multi-operator networks for mass market communication that is standardized, well defined, performance-optimized global service domain and focused on service delivery across the operators [23]. The IMS allows the solution to provide new multimedia rich communications by mixing telecommunications ad data on an access independent IP based architecture as defined by 3rd Generation Partnership Project (3GPP), 3rd Generation Partnership Project 2 (3GPP2) and Internet Engineering Task Force (IETF) standards. Network controlled multimedia services are then offered by the service providers using the NGN based IMS networks. The internet does provide some of the services that IMS will offer i.e. Voice over IP (VoIP) but the internet services are best effort services. The IMS provides Voice and Video over the managed IP network hence referred to as Voice/Video over Broadband (VoBB) and it does provide roaming facilities as well. For IMS to achieve its goals, it provides the support of peer-to-peer IP communications between existing technologies and also provides the framework for inter-operability of voice and data for both fixed and mobile networks. 3GPP R7 standards has been formulated to provide voice and video call continuity between different access technology making use of the IMS platform i.e. Call handing over from CDMA to GSM network.

The control infrastructure in the Circuit switched networks is replaced by IMS and it also separates the services from the underlying networks that are carrying them. In IMS, the signaling and media plane are separated which is not the case for PSTN. The signaling plane is used to handle session control, authentication and QoS aspects while the media encoding and transport issues are handled by the media plane. The separation of planes is not new in the telecommunication era as it started by the decomposition of MSC into a Media Gateway (MGW) and Mobile Softswitch or MSC server (MSS) when the mobile networks evolved form 3GPP R99 to 3GPP R4 so IMS is the continuation of that plane separation.

The convergence of multimedia services with fixed and mobile networks has created new demands for the high quality and user responsive service provision management [24]. The IMS makes use of Session Initiation Protocol (SIP) as the basic common signaling protocol in the platform. SIP is used for creating, modifying and terminating sessions in the IMS platform. These sessions include internet telephone calls, multimedia distribution and multimedia conferences

[25]. SIP enables internet endpoints which are also called user end points in order to create a session between users. SIP is a request-response protocol which has a resemblance to the two other Internet protocols, HyperText Transfer Protocol (HTTP) and Simple Mail Transfer Protocol (SMTP) which are the protocols used for Internet and e-mail respectively and therefore SIP sits alongside internet applications. The use of SIP enables telephony to become another web application and be able to be integrated to other internet services.

Motivation

The new technology innovations have given birth to new multimedia applications like VoIP, IPTV and Presence services. The convergence has provided an increase in the demand for services which have performance characteristics which is the proposition which defines the future of the service providers and network operators. There are some advantages in providing the IP multimedia services as opposed to the basic IP connectivity from the subscriber, network operator and third party application vendors i.e. ISP's. When subscribers are using basic IP connectivity which is provided by the ISP's, they are compelled to use the third-party providers for the IP multimedia services and applications.

As there is a significant growth of basic IP connectivity services like VoIP services provided by Skype, the network operators may lose potential revenue by not providing the VoIP call capability on their networks. Since the VoIP services provided by providers like Skype are the best effort services it is very important for the network operators to provide these services over managed IP network therefore providing Quality of Service (QoS), Reliability and prioritization i.e. voice packet prioritization.

Network operators have currently deployed the 2G (GSM) , 3G (WCDMA), CDMA, PSTN, WiMAX networks but the idea of convergence have changed the mindset with the development of a business model that provides the quick introduction of new services, exploit the potential revenue from these new services and have the flexibility of scaling the network. Network operators should focus on the management of network resources in order to optimize the performance of the services that they are delivering.

The customers now need the network operators to be able to provide services that have tailor made QoS characteristics according to the customer needs. Operators should then manage and monitor the network traffic in order to improve its performance and be able to deliver the agreed Service Level Agreement (SLA) which has been promised to the customers. Effective monitoring of the network performance and analysis of the network data to correlate end-to-end service performance is essential for the thorough understanding of network behavior [26].

The deployment of new applications by network operators creates huge demands on the network infrastructure as well as network operations staff. There is a need to provide scalable and flexible network which will make it possible to integrate diverse applications, reduce complexity and reduce Operational Expenditure (OPEX) of the network.

The challenges for the operators to provide these services in today's highly competitive market are:

- Huge potential customer growth.
- Seamless transition from the current CS networks to next generation networks.

The IMS provides a solution for seamless transition of the networks as well as its huge capacity and compactness makes it the network architecture to move forward. The IMS is the technology that has all the capabilities of driving the current networks to the future but it comes with a burden when it comes to signaling traffic. The IMS generate more signaling traffic compared with the current SS7 networks. This is due to the fact that IMS uses SIP as its core protocol. SIP is a text based protocol and the SIP requests can be bulky hence causing the signaling overload on the IMS networks.

The research focuses on the methodologies to optimize the IMS platform from the viewpoint of network operators. The signaling traffic generated by the IMS is characterized and analyzed and compression of SIP is done by using the text based compression methods. The queuing model of the IMS has been analyzed as well and using the analysis is done mathematically to evaluate the use of Hyper-threading technology in order to improve the performance of the IMS elements.

1.1 Key Research Questions

How can we improve the performance on IP Multimedia Subsystem?

- How can SIP be optimized?
- What Optimization methods can be used on IMS Elements?

1.2 Approach and Methodology

The aim of this research is to apply the optimization methods on the SIP and IMS elements to improve the performance on the IP Multimedia Subsystem.

A detailed study of the IMS has been undertaken. The study has focused on the layered architecture of the platform, the elements of the IMS and their functions as well as the reference interfaces of the IMS network.

The signaling generated on the IMS has been studied with the main focus on SIP as it the core protocol being used on the IMS platform. Other signaling protocols like H.248, SDP and Diameter have been carefully studied.

SIP is a text based protocol so it is important to take this into an advantage. The concept of Intelligent Compression of SIP is going to be used to realize the reduction of signaling traffic in IMS. Intelligent compression is also termed as loss-less compression and is used on text file compression. I will apply this technique on SIP messages and making use of Burrows-Wheeler Transform (BWT) and Move to Front Encoding algorithm. The C++ programming has been used to test the BWT compression algorithms.

Having studied the elements on the IMS network the queuing analysis of the IMS has been formulated. The impact of the use of Hyper-Threading technology on the IMS elements has been studied and the results have been demonstrated. The queuing models (M/M/1 and M/M/2) have been used be used to perform the mathematical analysis. Octave Coding has been used to provide the comparative graphical analysis for the queuing analysis when Hyper-threading was used on the IMS elements.

1.3 Report Layout

Chapter 02

Chapter 02 is the literature review chapter. It details the study of the publications that have been done on the topic being researched. The focus is the optimization of SIP as well as the optimization of IMS elements in order to improve the performance of the platform.

Chapter 03

Chapter 03 outlines the analysis of IMS platform. The layered architecture of the platform, the elements as well as the reference interface as on the IMS platform is outlined.

Chapter 04

This chapter outlines the characterization of the signaling traffic on the IMS platform. The protocols studied include the SDP, Diameter, H.248 and the main signaling protocol for the IMS which is SIP.

Chapter 05

In this chapter the detailed analysis of SIP signaling compression is achieved. The BWT is used to perform the compression and the results are tabulated and displayed graphically.

Chapter 06

Chapter 06 details the queuing theory used in the IMS as well the use of Hyper-threading technology on the IMS elements in order to improve the performance.

Chapter 07

The conclusion and future work is outlined. The chapter shows that the research questions could be answered and importance of this work towards the converged networks.

Chapter 02
Literature Review

2. Literature Review

The birth of SIP as the signaling protocol has created a lot of interest in the researchers and academics. As SIP is text based the request messages are quite bulky and they are generating extensive signaling overload. The review of the publications on how the signaling traffic can be reduced is outlined in this literature.
The IMS elements need to be optimized as well in order to ensure that the performance of IMS is enhanced.

2.1. SIP Load Balancing

2.1.1 Message Based Load Balancing (MBLB)

The use of message-based load balancing offers the ability to aggregate and disaggregate SIP messaging, and prevents message-blocking asymmetric server return [27]. Aggregation is used to combine multiple communication streams within a single stream connection. SIP is affected by TCP or UDP port as well as IP triplet combinations at the service provider level. Disaggregation is used to separate individual messages out of a single, shared TCP connection

The implementation of aggregation and disaggregation on IMS improves user experience, increase performance, decrease the bottlenecks in the platform and improves the session setup time.

Since SIP is using message-based context, the ability to perform load balancing of messages is called Message Based Load Balancing (MBLB). When SIP is used for voice connection it is important the session be maintained for the duration of the call and this requires the connection between the SIP server and the client device to be properly maintained to avoid the failure of the communication channel. When the data channel experience service degradation, the control channel will pass a stream of messages to re-establish the required quality of service and this directly affects the user experience. MBLB is used to provide a message-by-message look within the SIP communications to determine and maintain the correct server connections [27].

2.1.2. SIP Load Balancer

The use of a SIP load balancer ensures that all network elements and servers can be utilized efficiently. Inbound/outbound traffic is handled in a timely and successful manner, avoiding hold ups, traps and other overwhelming processes that may cause bottlenecks or crashes [28]. The IMS network does not only need to have an efficient load balancer, but it must also optimize the SIP server platforms and the entire IMS elements. When a server platform crashes or comes into a bottleneck, data traffic can be exposed to harmful security violations such as Denial of service (DoS) or Distributed Denial of Service (DDoS) attacks [28]. The combination message load balancing, SIP optimization and security therefore enables the IMS to be a highly efficient and scalable system.

2.2. SIP Compression

SIP-based communications use the combination of TCP and UDP-transported data for the signal control and the exchange of data. The separate streams need to be correlated and used together to provide an efficient communication for the users that are participating in SIP-based dialogs.

The implementation of Session Border Controllers (SBCs) at the edge the core and access network provides a number of functions that are related to SIP and its flexible nature. The reliability and scalability of SBC ensures the successful deployments of SIP because the SBC provides much of the functionality that makes IP-based telephony services work efficiently.

SIP can have the benefits of compression as it is a text based protocol. Offloading compression from SIP servers and SBCs to a high-availability solution affords greater benefits in the performance of the entire infrastructure, as most high-availability solutions provide hardware-based compression [29].

The other form of compression takes advantage of SIP's mechanism for representing common header field names in an abbreviated form. This provides a useful solution when messages become too large to be carried on the available transport network, i.e. when the Maximum Transmission Unit (MTU) is exceeded on using UDP. The abbreviated form of a header can be substituted at any time, and can appear in both long and short form within the same message [29].

This header field name variability is used to provide information to high-availability (HA) solutions and it can cause problems if the HA solution cannot recognize the different variations. It is either necessary for the HA solution to be capable of recognizing either forms of the headers or to provide the means by which a mapping can be created, this is to ensure that continuity is not compromised.

2.3. Co-Locating of IMS servers

The IMS is an architecture where for a typical call, five Call Session Control Function (CSCF) proxies must be interrogated as well as several application servers in order to provide a particular service to the subscriber.

When IMS servers like Proxy CSCF [P-CSCF] and Serving CSCF [S-CSCF] are running on the same host and are using an optimized architecture, the SIP message can be internally passed between SIP servers using a simple and very light operating system (OS) mechanism, like a queue. In that case, the SIP message never leaves the host, and the IP network is not used at all [30]. This arrangement provides an improvement in Capacity, Performance and Reliability of the IMS Network.

2.3.1. Capacity Improvement

Capacity improvement is achieved by the reduction of service processing time. When the two SIP servers are running on the same host, layers 1 to 5 on the SIP stack are skipped causing a significant decrease in processing that is required to treat each SIP transaction. These layers are the Physical Layer, Operating System – Interrupt service Routine (ISR), Internet Protocol (IP), Over IP transport (UDP-TCP), Serialization and Parsing. Parsing alone utilizes 25% of the processing time and skipping this step provides a great deal in reducing processing time. When two SIP servers are running on the same host and an optimization technique is used to ensure that the 5 layers are suppressed, there is a 45% saving on the processing time when the SIP messages are exchanged internally between two IMS Servers [30].

2.3.2. Performance Improvement

The performance improvement on the co-located IMS servers is measured by the SIP latency which is the time spent when a SIP server sends a message to another server to when the second server receives the message. The latency of IMS messages between two servers is about 25ms at nominal load. Having the IMS servers on one host we can have the latency which is less than 1ms.

2.3.3. Reliability Improvements

The reliability of IMS is inversely proportional to the number of SIP servers [30], which are required to make a successful call. This is because SIP servers are the potential points of failures in the network. Even though SIP servers have redundant hardware, there is still a probability of the entire server to fail. When several IMS servers in the network that are used for the same call run on one host, the call will transverse fewer hardware elements and therefore there are fewer points of failure. This will increase the reliability of the network.

2.3.4. Implementation Techniques for Co-locating IMS servers

There is a need for a network to be engineered and provisioned in such a way that IMS servers are effectively co-located when they are serving the same call. The software architecture need to allow IMS servers i.e. P-CSCF, S-CSCF, I-CSCF and Breakout Gateway Control Function (BGCF) as well as Application servers to run on the same host and have effective communication between each other, making full use of co-location.

Co-locating the IMS servers on the same host will improve the overall efficiency of the network without having to do alterations on the architecture itself.

Chapter 03
Analysis of IP Multimedia Subsystem

3. Background of The IP Multimedia Subsystem

The IP Multimedia subsystem was initially defined by the 3GPP and 3GPP2 wireless working bodies [2]. The focus of IMS is to provide a new network architecture to enable the convergence of voice, data and multimedia services over an IP based network infrastructure. Multimedia services include real-time communications like speech, audio, video and text applications. The information management system has been designed to reduce the divide between the traditional telecommunication networks and the internet technology. The services that the IMS has to provide are already available to many subscribers: The services include rich voice services, video telephony, conferencing and Instant Messaging (IM). These services are currently provided as best effort services. In order for the IMS to support the sophisticated demand of these services there has to be some key mechanisms to be followed. These mechanisms include session negotiation and management, Quality of Service (QoS) and mobility management [2]. The IMS has the ability of providing IP interoperability of real time services between fixed and mobile networks and this is the basis of seamless converged voice and data services.

3.1 Signaling in IP Networks

In the fixed-line networks (PSTN) the IP signaling has already been deployed. The IP signaling protocol in PSTN is H.323 which is used when a softswitch is used on the network and is used to carry class 4 and 5 traffic types.
The IMS is developed around the Session Initiation Protocol (SIP). SIP provides some advantages over its predecessor the H.323. SIP is an application layer protocol that is used to establish, modify and terminate multimedia sessions over an IP based network.

3.2 Layered Architecture of IMS

The definition of IMS by 3GPP is an all packet core network that is able to accept all types of access networks i.e. WiMAX, CDMA, GSM in order to deliver a wide range a multimedia services which can be offered to the user by any device connected to the network. The use of SIP in IMS enable the support of IP-to-IP sessions over any wire-line connection systems like DSL as well as wireless networks like Wi-Fi, GSM and CDMA. The IMS allows the inter-working between the traditional TDM networks and the IP networks.

3.2.1 Access Layer

The IMS architecture is independent of any access bearer. In the mobile networks the access layer could be any or a combination of the following: General Packet Radio Service (GPRS), Enhance Data Rates for GSM Evolution (EDGE) , Code Division Multiple Access (CDMA), Wireless

Interoperability for Microwave Access (WiMAX), Universal Mobile Telecommunication System (UMTS) and Wireless Local Area Networks (W-LAN or Wi-Fi).
The fixed line networks make use of Asymmetric Digital Subscriber Line (ADSL) and cable network accesses.

3.2.2 Transport Layer

This is an all IP network which consists of IP Routers. These routers are Label Edge Routers and Core switching routers. The IP/MPLS (Multi-protocol label Switching) is the transport layer technology for the IMS platform. MPLS defines a mechanism for the forwarding of packets in a router network. Due to its flexibility, the MPLS has become the default IP transport network for the Next Generation Networks making use of the IMS core in order to provide reliability and Quality of service.

3.2.3 Session Control Layer

This layer consists of network control servers which are used for the management of calls, establishment and modifications of sessions in the IMS platform. The two main elements in this layer are the Call Session Control Function (CSCF) and the Home Subscriber Server (HSS) [3]. These two elements form the core of the IMS architecture and are sometimes referred to as SIP servers. The CSCF provides end-point registration and routing of SIP signaling messages [3] and provides interworking with the transport layer for the guaranteed QoS of all services. The HSS is the database that is used to store subscriber service profiles and service triggers. The other information stored by the HSS includes the dynamic data of subscribes like location information.

3.2.4 The Application Layer

This layer makes use of application and content servers to provide value-added services. The Application Servers (AS), the Multimedia Resource Function Controller (MRFC) and Multimedia Function Resource Processor (MRFP) form the core of this of this layer. The AS is responsible for the execution of service-specific logic i.e. call flows and user interactions with subscribers [3]. The MRFP is also known as the IP media server is used to provide media processing for the application layer. The media server is used to enable the delivery of some non-telephony services like push-to-talk (PTT), speech enabled services [3], video services and other services like conferencing, prepaid cards and personalized ring-back tones [3].
The control and application layers are access and transport independent [3] and this is very useful in order to ensure that the user is able access the IMS services required from any access network and from any connection device.

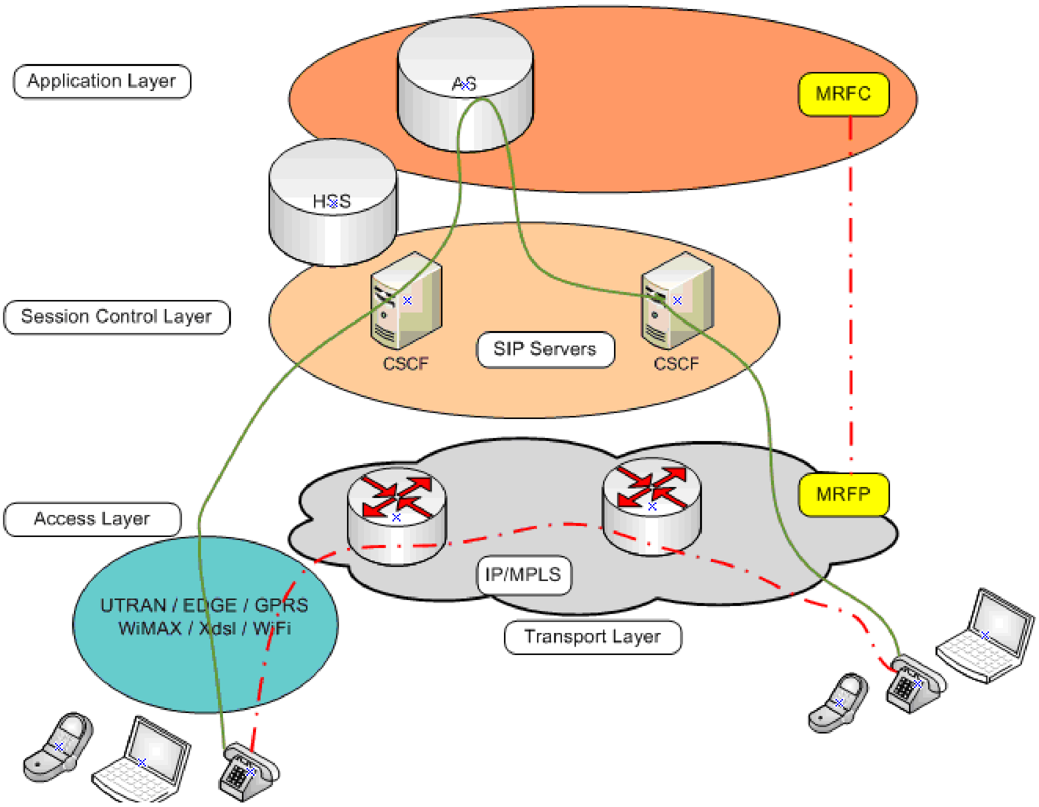

Figure 01 : Layered Architecture of IMS

3.4. The IMS Architecture

Following the idea of service developers to move away from vertically integrated service silos, the IMS architecture is used to provide services that are horizontally integrated. The IMS provides common functionality that is specific to a particular service [4] and in that way the service developers are only needed to implement particular service specific functionality. Functions provided by the IMS that are available to services built on top of it include capability negotiation, authentication, service invocation, addressing, routing, group management, presence, provisioning, session establishment and charging [4]. Since the IMS provides the creation of services in a speedy manner it is referred to as service enabling architecture.

The standardization of IMS by 3GPP allows the definition of logical functions, not nodes. Several ways are used to map the functions to physical nodes. A single node can be used to implement a single function or several functions and a single logical function can be implemented into several IMS physical nodes.

3.4.1 Call Session Control Function (CSCF)

The main functional entity in the IMS is the CSCF [1]. The CSCF is mainly a SIP server but it also implements non-SIP interfaces which are Diameter-based interfaces. The CSCF is further decomposed into three functional entities; Proxy-CSCF (P-CSCF), Serving-CSCF (S-CSCF), and Interrogating-CSCF, I-CSCF.

Serving-CSCF

The S-CSCF is used for registration, session control, as a proxy server and forms interactions with application servers. The other functions performed by the S-CSCF include the service-related event notifications to the users [1], as well as generating Call Data Records (CDR's) which are needed for accounting and billing purposes.

Registration

The S-CSCF acts as a SIP Registrar [1] [4], and is used to accept the registrations from the users as well as location information from the location servers such as Home Subscriber Server (HSS). This is done by the implementation of Cx interface towards the HSS. The Cx interface is Diameter-based signaling protocol.

Session Control

The S-CSCF performs SIP session control functions for a registered user [1] as well as relaying the SIP requests and responses between the parties in call.

Proxy Server

The S-CSCF can also act as SIP Proxy server and is used to relay SIP messages between the users in the call as well as other SIP servers.

Interactions with other Application Servers

An S-CSCF acts as the interface to the Application Servers (AS) and other IP or legacy service platforms [1]. A user's filter criteria are a collection of triggers that are applied to user's incoming and outgoing SIP Messages [4]. When a trigger matches a message that message is then relayed to the AS that is indicated in the trigger.

Proxy-CSCF

The P-CSCF is the first contact point inside a local or a visited IMS [1] and can be referred as a SIP outbound proxy. All the traffic from the user terminal passes through the P-CSCF. The Policy Control Function (PCF) is contained in the P-CSCF and is used to control the policy on how the GGSN bearers should be used.

The Specific functions performed by the P-CSCF include:

- Forwarding of SIP REGISTER requests which are from the terminal's home network.
- SIP Message forwarding from a user terminal to the SIP server and back
- Perform necessary modifications to the SIP Requests before forwarding them to the other entities [1].
- Maintenance of security associated with the user terminal as well the generation of Call Data Records.

Interrogating-CSCF (I-CSCF)
The I-CSCF is an optional function that can be used to hide an operator network's internal structure from an external network [1]. This server act as a central contact point in the network and is used for all sessions that are intended for a subscriber of a particular network or the visiting user that is currently roaming on the network. It is used primary to select the S-CSCF for a user's session [1], routing of SIP Requests and generation of Call Data Records.
The following criteria is used by the I-CSCF in order to select the S-CSCF:

- Capability requirements by the user
- The S-CSCF's capabilities and availabilities
- Topological information, such as the location of the S-CSCF and the location of user's P-CSCF if they are in the same operator's network as the S-CSCF [1].

3.4.2 Home Subscriber Server (HSS)

The HSS is one of the key concepts of the converged IMS network solution. This is a database that is used to store the user profiles in the domain. The user profiles include security related information such as cryptographic keys [4] and all the service-related information which the user has subscribed to. As the networks grow there need to be more than one HSS in the networks and when this happens these networks will also implement the Subscriber Location Function (SLF). Nodes that need to query the HSS about a particular user first query the SLF, which returns the address of the HSS handling the user [4]. The interfaces on the HSS and SLF make use of Diameter-based interface. The interfaces implemented by the HSS and SLF are Cx and Dx respectively. The SLF can be effectively seen as a Diameter redirection agent.

3.4.3 PSTN / GSM Interworking Functions

There is a need for IMS networks to be able to interact with PSTN in order for the IMS user to establish services the PSTN users. PSTN and legacy networks perform the interworking with IMS at two levels:

- User plane that is used for IP Media stream conversion on the IMS side to TDM media streams that are suitable for the PSTN
- The signaling plane that is used to covert SIP signaling form the IMS to the legacy signaling such as ISUP (ISDN User Part)

The IMS-Media Gateway (IMS-GW) is responsible for user plane functions on the IMS platform. The MGW performs the translation between the RTP-based media used in the IMS and the media format used in the circuit switched networks [4].

The Media Gateway Control Function (MGCF) is used to perform the signaling plane functions i.e. the protocol conversion between SIP and ISUP. The MGCF is also to control the Media Gateway through the H.248-based Mn interface. Signaling gateway is used in order to form the interconnection between the MGCF and the legacy PSTN networks.

The Breakout Gateway Control Function (BGCF) decides which MGCF handles a particular session that will terminate in a circuit switched network [4]. The BGCF has also the responsibility of forwarding the session signaling to the desired MGCF as well the BGCF in the destination circuit switched network or PSTN.

3.4.4 The Application Servers

Application servers are used for the implementation of IMS services. The application servers form the communication with S-CSCF and HSS through the SIP based ICS interface and Diameter based Sh Interface respectively. The new services on the IMS environment are mainly implemented in the application servers but it is also crucial to be able to provide legacy services to IMS users. The IMS specifies the Open Service Access (OSA) – Service Capability Sever to interwork with the OSA framework application server and the IP Multimedia Service Switching Function (IM-SSF) to provide Customized Applications for Mobile networks Enhanced Logic (CAMEL) services [4]. The HSS is accessed by the Service Capability Server by the Diameter based Sh interface and Service Switching Function by Diameter based Si Interface. The communication between these servers and the S-CSCF is achieved by the ICS interface which is based on SIP.

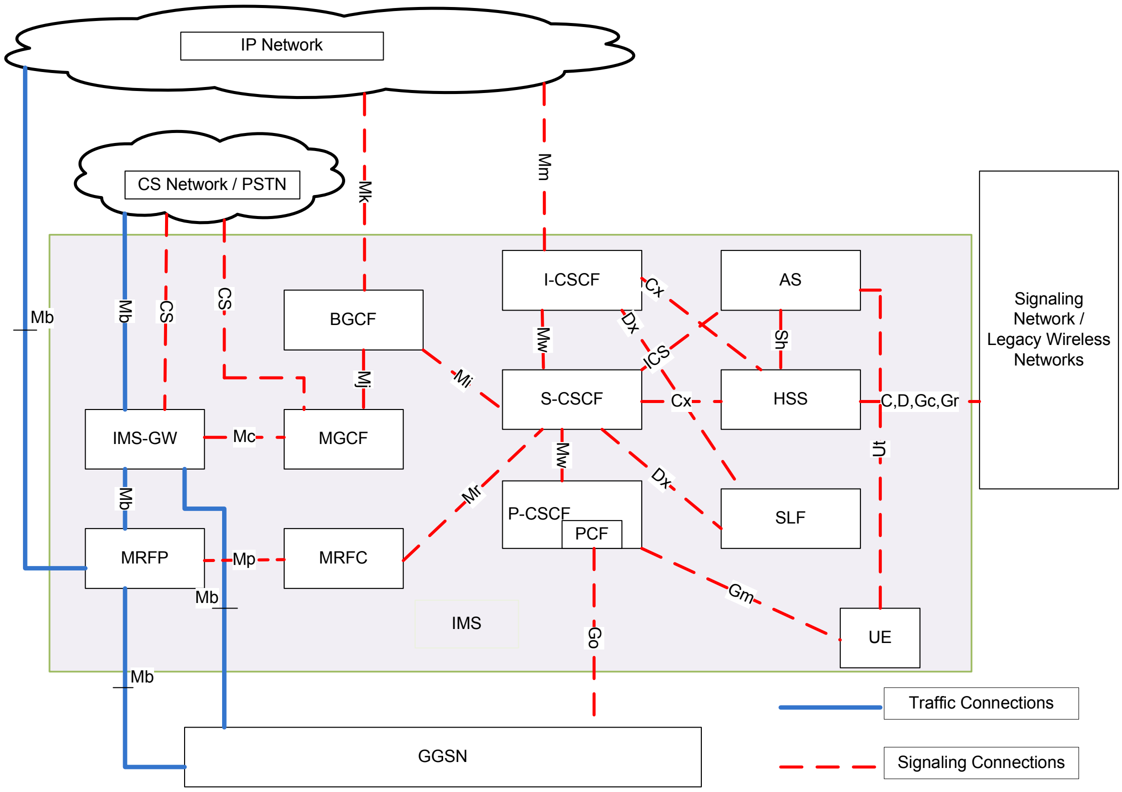

Figure 02 : IMS Architecture

3.5 Reference Interfaces in IMS

Interface	Interacting Entities	Protocol	Function
Gm	UE ←→ P-CSCF	SIP	Used to carry SIP Signaling information between the UE and P-CSCF
Ut	UE ←→ AS	XCAP	This is the OMA-defined interface which is used

			for the UE application clients for the management of service related data which is maintained by the application servers
Go	GGSN ←→ PCF	COPS	This is an optional interface which is used between GGSN and PDF for initialization procedure.
Mb	MRFP ←→ UE IMS-GW ←→ UE	RTP	This is the IP interface between the MRFP or the IMS-GW and the UE through the GGSN which carries the IP packets which are corresponding to media streams. This interface is also used between IMS gateway and the PSTN networks
Mc/Mn	IMS-GW ←→ MGCF	H.248/Megaco	This interface is used for the control of IMS-GW by the MGCF
Mp	MRFC ←→ MRFP	H.248/Megaco	Media Resources of the MRFP are controlled by the MRFC using this interface
Mj	BGCF ←→ MGCF	SIP	The Mj interface is used when the BGCF and MGCF are on the same IMS core network.
Mk	BGCF ←→ BGCF	SIP	This interface is used for the interaction of two BGCF's from different IMS networks
Mi	S-CSCF←→ BGCF	SIP	The Mi Interface is used for the transportation of signaling information from S-CSCF to the other CS networks using a breakout functionality
Mr	S-CSCF←→ MRFC	SIP	When the media resources are requested by the S-CSCF from the MRFC this interface is then used to provide support.
Mw	P-CSCF←→ S-CSCF P-CSCF←→ I-CSCF	SIP	This interface is used between CSCF's for