IMPLEMENTING SECURITY IN AN IP MULTIMEDIA SUBSYSTEM (IMS) NEXT GENERATION NETWORK – A CASE STUDY

By

Jose M. Ortiz-Villajos

A Thesis Submitted to the Faculty of

The College of Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, Florida

April 2009

IMPLEMENTING SECURITY IN AN IP MULTIMEDIA SUBSYSTEM (IMS)
NEXT GENERATION NETWORK – A CASE STUDY

By

Jose M. Ortiz-Villajos

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Eduardo Fernández, Department of Computer Science and Engineering, and has been approved by the members of his supervisory committee. It was submitted to the faculty of the College of Engineering and Computer Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Eduardo Fernandez, Ph.D.
Thesis Advisor

Michael VanHilst, Ph.D.

for Jie Wu

Jie Wu, Ph.D.

Robert B Cooper

Borko Fuhrt, Ph.D.
Chair, Department of Computer Science and Engineering

Karl K. Stevens Ph.D., P.E.
Dean, College of Engineering and Computer Science

Barry T. Rosson

Barry T. Rosson, Ph.D.
Dean, Graduate College

April 9, 2009

Date

ABSTRACT

Author:	Jose M. Ortiz-Villajos
Title:	Implementing Security in an IP Multimedia Subsystem (IMS) Next Generation Network – A Case Study
Institution:	Florida Atlantic University
Dissertation Advisor:	Dr. Eduardo Fernandez
Degree:	Master of Science
Year:	2009

The IP Multimedia Subsystem (IMS) has gone from just a step in the evolution of the GSM cellular architecture control core, to being the de-facto framework for Next Generation Network (NGN) implementations and deployments by operators world-wide, not only cellular mobile communications operators, but also fixed line, cable television, and alternative operators. With this transition from standards documents to the real world, engineers in these new multimedia communications companies need to face the task of making these new networks secure against threats and real attacks that were not a part of the previous generation of networks. We present the IMS and other competing frameworks, we analyze the security issues, we present the topic of Security Patterns, we introduce several new patterns, including the basis for a Generic Network pattern, and we apply these concepts to designing a security architecture for a fictitious 3G operator using IMS for the control core.

IMPLEMENTING SECURITY IN AN IP MULTIMEDIA SUBSYSTEM (IMS)

NEXT GENERATION NETWORK – A CASE STUDY

TABLES

FIGURES

GLOSSARY

Second Generation (2G) Cellular Network: the term used to describe the migration to all-digital cellular mobile communications. Even with the possibility of transmitting data afforded by the increased transport capacity of 2G, the service offering available is mostly voice. The two most widely deployed 2G technologies are GSM and CDMA/IS-95.

Third Generation (3G) Cellular Network: the group of technologies and standards that together bring about the possibility of much higher data rates in cellular communications than previously available. Depending on the environment (fixed, pedestrian or vehicular), data rates from 144Kbps to 2Mbps are possible. 3G is being used today for both voice and data offerings.

Third Generation Partnership Project (3GPP): a consortium of standards organizations and market representatives formed in December 1988 with the purpose co-operating for the production of globally applicable Technical Specifications for a 3rd Generation and beyond Mobile System.

Fourth Generation Cellular Network: The term used to group the different technologies which will improve on 3G. Among them, WiMax and LTE (Long Term Evolution) are the most prominent.

Access Border Gateway Function: network element at the edge of the operator’s boundary towards the access network where the media traverses from one domain to the other.

Authentication and Key Agreement (AKA): mobile network protocol for mutual authentication between mobile terminal and operator network and for exchange of integrity and cryptography keys.

Application Server: in the IMS architecture these are the servers within the application layer which provide end-user services such as voice, messaging, gaming, etc.

Back to Back User Agent **(B2BUA):** network entity where two back to back SIP clients are connected via session processing logic such that the session legs are independent of each other.

Breakout Gateway Control Function (BGCF): IMS logical entity for handing sessions off to the PSTN/PLMN**.**

CableLabs: the research and development consortium founded in 1988 by cable operating companies to develop new cable communications technologies.

CALEA: name of the U. S. wiretapping law passed in 1994 which “*makes clear a telecommunications carrier's duty to cooperate in the interception of communications for Law Enforcement purposes”*.

Call Session Control Function (CSCF): IMS logical entity for session layer handling of voice and multimedia connections and which uses SIP as session protocol.

De-Militarized Zone (DMZ): an area in between two firewalls where a server or proxy is placed which relays traffic between the secure and the non-secure domains of an operator’s network.

Data over Cable Service Interface Specification (DOCSIS): the CableLabs specification that standardizes how to transport data to and from end-customers over cable television networks.

Emergency Call Session Control Server (E-CSCF): the class of CSCF introduced in 3GPP R7 explicitly for handling originating emergency calls (911).

Emergency Services: the term used in telecommunications to denote the components, protocols, and functions related to the processing of 911 calls.

ETSI: the organization of companies and individual members which seeks to produce telecommunication standards to be used in Europe.

Fully Qualified Domain Name (FQDN): a unique domain name within the Domain Name System's tree hierarchy.

GPRS Gateway Switching Node (GGSN): a network node in the 2G GSM packet network that acts as gateway between the GPRS and other networks.

GPRS: a packet data service for users of 2G GSM cellular networks which provides data rates of 56-114 Kbps.

Global Switching Mobile (GSM): a second generation (2G) digital cellular telephony standard. In the U.S. it is the technology used by T-Mobile and AT&T.

Home Location Register (HLR): in a GSM network, it is the central database which contains the subscriber information for services and registration location.

Home Subscriber Server (HSS): in an IMS network, it is the central database which contains the subscriber data, services to which the subscriber has access, as well as authentication information.

Interworking Border Control Function (I-BCF): in an IMS network, it is the logical entity which is placed at the logical border of the operator's network to control what signaling enters and exits the network, as well as to control the media flowing via the IBGF.

Interworking Border Gateway Function (I-BGF): in an IMS network, it is the logical entity through which all media exits and enters the network to and from other IP networks. It may implement firewall and policy control functionality.

Interrogating Call Session Control Function (I-CSCF): in an IMS, the logical entity which interrogates the HSS upon user registration or session initiation form a partner network, to find out which S-CSCF should be responsible for the subscriber or session.

Internet Engineering Task Force (IETF): an open international community of network designers, operators, vendors, and researchers, which develops and promotes Internet standards.

IP Multimedia Subsystem (IMS): an architectural framework for delivering communications voice and multimedia services over Internet Protocol networks. Originally developed by the 3GPP as an evolution of GSM networks for cellular services, it now addresses services over any type of access.

Location and Presence: a communications networks logical entity which collects and provides information on a subscriber's registration and availability status, as well as his or her location.

Lawful Interception: the lawfully mandated collection of call content (media) and envelope information (signaling) pertaining a voice or other type of electronic communication.

Long Term Evolution (LTE): the term used by 3GPP to denote the next step (after 3G) in cellular wireless communications.

Media Gateway Control Function (MGCF): in an IMS network the logical entity which controls the media gateway through which IP sessions are converted into TDM traffic towards the PSTN or PLMN. It usually also is in charge of the SS7 signaling towards the PSTN or PLMN network.

Media Gateway Control Protocol (MGCP): the protocol used between the Media Gateway Control Function and the Media Gateway.

Media Gateway (MGW): the entity which converts media from IP into TDM and vice-versa, between an IP and a TDM network.

Media Resource Function (MRF): a network element in an IMS or IP network which is responsible for playing announcements, tones, and other media streams towards subscribers, and which can also receive and interpret certain media, for example DTMF tones.

Mobile Switching Center (MSC): in a cellular network this is the network element responsible for switching calls, providing call features, signaling to other networks, handling mobility, and other functions of cellular phone service.

Multi-Services Operators (MSO): communications companies primarily dedicated to providing cable television service to consumers, but also broadband internet and voice over IP phone services.

Network Address Translation (NAT): network element situated in between a subscriber's home IP equipment and the operator's network, which translates between the private IP and port addressing in the subscriber's domain, and the public IP domain on the other side.

Next Generation Network (NGN): a communications network breaking with traditional telephony networks in that it generally: is based on Internet Protocol, effects separation of signaling and media, separates switching from applications, and is based on open interfaces and commercial IT hardware instead of proprietary platforms.

Open Mobile Alliance (OMA): an industry forum founded in 2002 by communications and information technology manufacturers and operators to pioneer and enable the development of end-to-end mobile services.

PacketCable: the initiative within CableLabs to develop the specifications for advanced multimedia communications over cable.

Private ID: in SIP, a unique identifier belonging to a user and device which takes the form specified in RFC 2486 and used for subscription identification and authentication purposes. It is typically associated with a particular device, e.g. bobs.pda@operator.net, or bobs.homephone@operator.net.

Public ID: in SIP, one or more identifiers allocated to a user used to route SIP signaling messages. It is typically associated with a user's desired service, e.g. bob.home@operator.net, or bob.business@operator.net.

Policy and Charging Rules Function (PCRF): in an IMS network the logical entity which administers and directs service and media policy, by matching requested service levels with agreed subscriptions and overall network conditions.

Proxy Call Session Control Function (P-CSCF): in an IMS network the Proxy CSCF is the SIP entity that interfaces directly with the SIP client in the user device. It is responsible for access security, for signaling compression if needed, and for forwarding all SIP messages to the correct Serving CSCF.

Public Land Mobile Network (PLMN): term regularly used to denote a legacy cellular network.

Public Switched Telephone Network (PSTN): term used to denote the legacy fixed telephone network.

Quality of Service (QoS): in voice over IP communications the term used to specify whether the network is actively influencing certain packet transmission characteristics (packet delay, packet loss, packet jitter) in order to give preferential treatment to certain packet streams at the expense of others.

Request For Comments (RFC): the technical documents produced by the IETF by which new protocols or extensions to existing protocols get defined. There are three levels of RFCs: proposed standard, draft standard, and Internet standard. Only a few reach the last stage.

Rich Communication Services (RCS): in an IMS network a set of standards that define how several individual services, such as voice calls, video calls, text messaging, multimedia messaging, network phone book, etc. interact together in a SIP client. They define the “look and feel” and guarantee that terminals and network equipment from different vendors will interoperate.

Session Border Controller (SBC): a network element which is placed at the border of the operator’s secure domain and which encompasses the functions of firewall and application layer gateway, with some others which are vendor specific and can include protocol conversion, policy control, and Proxy-CSCF.

Serving Call Session Control Function (S-CSCF): in an IMS network, the S-CSCF is the session-stateful SIP server which controls subscriber authentication, holds the registration status of every subscriber, which determines session routing, and which invokes services from the application layer.

Serving GPRS Switching Node (SGSN): the routing network element within a GPRS network which performs mobility management, security and access control functions for data connections.

Subscriber Identity Module (SIM): the SIM is a logical module within the secure chip in a GSM phone which contains vital subscriber information and performs the computations necessary to derive cryptographic keys for authentication, confidentiality and integrity.

Session Initiation Protocol (SIP): a text-based signaling protocol derived from HTTP which is used to initiate, modify, and terminate communications sessions among two or more clients. It is the chosen session layer protocol for 3GPP IMS.

Service Level Agreement (SLA): a contract between two operators which formalizes the agreements for handling the traffic between them. It may include limits on number of calls accepted, bandwidth used, billing, quality of service provided, etc.

Softwswitch: a term used in next generation networks architecture to denote a communications node which unlike in previous technologies, only performs signaling switching, media control, and call processing functions, without actually processing the media itself.

TISPAN: a standardization body created in 2003 by ETSI for the purpose of defining and producing the specifications for the Next Generation Network.

Time Division Multiplex (TDM): the technology used to transmit multiple Pulse Code Modulation (PCM) streams via the same physical wire or fiber connection by interleaving slower data rate digitized transmissions within a higher (by factor of “n”) rate stream, separated by timing pulses.

Transport Layer Security (TLS): a protocol, successor to SSL which is implemented in most Internet browsers, designed to provide data integrity and confidentiality in TCP communications. TLS is specified in RFC 2246.

User Agent (UA): a software entity which can take part in a higher layer communications session. It interfaces on one side with the application in the user device, and on the other, with the network. In the case of IMS, it uses SIP towards the network.

Universal Integrated Circuit Card (UICC): the removable smart card found in GSM phones and used to store a subscriber’s subscription information, authentication keys, phonebook, and messages. It contains the SIM and USIM logical applications.

Universal Mobile Telecommunications Services (UMTS): – The third generation (3G) cellular network for GSM.

UMTS Subscriber Identity Module (USIM): the logical module within a UICC, standardized in 3GPP TS 31.102, which provides subscriber parameters for authentication and other functions in 3G services.

UMTS Terrestrial Radio Access Network (UTRAN): the components of the radio access part of a 3G cellular network.

Voice Application Server (VAS): the network element in an IMS architecture which provides telephony services and features.

Visiting Location Register (VLR): the network element in a second generation network which contains subscriber-related data for a subscriber which is roaming within its coverage area, and only until it leaves the coverage area.

Voice over IP (VoIP): the technology and architecture which enables voice sessions over Internet protocol, as opposed to over traditional Time Division Multiplex.

Virtual Private Network (VPN): a Permanent encrypted layer-4 connection between two end points.

ACRONYMS

2G	Second Generation Cellular Network
3G	Third Generation Cellular Network
3GPP	Third Generation Partnership Project
4G	Fourth Generation Cellular Network
AAA	Authentication Authorization and Accounting
A-BGF	Access Border Gateway Function
AKA	Authentication and Key Agreement
AS	Application Server
AuC	Authentication Center
AVP	Attribute-Value Pairs
B2BUA	Back to Back User Agent
BGCF	Breakout Gateway Control Function
CALEA	Communications Assistance for Law Enforcement Act
CDMA	Code Division Multiple Access
CMTS	Cable Modem Termination System
COTS	Commercial Off-The-Shelf
CSCF	Call Session Control Function
DBS	Data Base Server
DMZ	De-Militarized Zone
DNS	Domain Name Server
DOCSIS	Data Over Cable Service Interface Specification
DoS	Denial Of Service
E-CSCF	Emergency Call Session Control Function

EMS	Element Management System
ENUM	Electronic Numbering
ETSI	European Telecommunications Standards Institute
FQDN	Fully Qualified Domain Name
FW	Firewall
GAA	Generic Authentication Architecture
GBA	Generic Bootstrap Architecture
GGSN	GPRS Gateway Switching Node
GPRS	General Packet Radio Service
GSM	Global Switching Mobile
HLR	Home Location Register
HSS	Home Subscriber Server
IBCF	Interworking Border Control Function
I-BGF	Interworking Border Gateway Function
I-CSCF	Interrogating Call Session Control Function
IDS	Intrusion Detection System
IETF	Internet Engineering Task Force
IM	Instant Messaging
IMS	IP Multimedia Subsystem
IPSec	IP Security
L&P	Location and Presence
LI	Lawful Interception
LRF	Location Resource Function
LTE	Long Term Evolution
MGCF	Media Gateway Control Function
MGCP	Media Gateway Control Protocol
MGW	Media Gateway
MRCP	Media Resource Control Processor

MRF	Media Resource Function
MSC	Mobile Switching Center
MSF	Multiservice Switching Forum
MSO	Multi-Services Operators
NASS	Network Attachment Subsystem
NAT	Network Address Translation
NDS	Network Domain System
NE	Network Element
NGN	Next Generation Network
NNI	Network to Network Interface
OMA	Open Mobile Alliance
OS	Operating System
PCM	Pulse Code Modulation
PCRF	Policy and Charging Rules Function
P-CSCF	Proxy Call Session Control Function
PDA	Personal Digital Assistant
PLMN	Public Land Mobile Network
PoC	Push to Talk Over Cellular
PSTN	Public Switched Telephone Network
PTT	Push To Talk
QoS	Quality of Service
RBAC	Role Based Access Control
RCS	Rich Communication Services
RFC	Request For Comments
RTP	Real Time Protocol
SBC	Session Border Controller
S-CSCF	Serving Call Session Control Function
SEG	Security Gateway

SGSN	Serving GPRS Switching Node
SGW	Security Gateway
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SMS	Short Message System
SNMP	Simple Network Management Protocol
SS7	Signaling System Number 7
TDM	Time Division Multiplex
TISPAN	Telecommunications and Internet converged Services and Protocols for Advanced Networking
TLS	Transport Layer Security
UA	User Agent
UICC	Universal Integrated Circuit Card
UMTS	Universal Mobile Telecommunications System
USIM	UMTS Subscriber Identity Module
UTRAN	UMTS Terrestrial Radio Access Network
VAS	Voice Application Server
VLR	Visiting Location Register
VoIP	Voice Over IP
VPN	Virtual Private Network

PART A

VOICE OVER IP AND MULTI-MEDIA NETWORKS

1 INTRODUCTION

1.1 Motivation

Security in voice telecommunications networks is not an area into which operators of such networks had to give a major share of their time and resources up to recently, and by recently we mean about the last 10 years or so. Granted that the topic has always been important ever since the means became available with which to obtain free services (I remember as a young boy using a modified piezoelectric gun of the kind used to light stoves, to “shock” a pay phone into believing I had deposited enough coins for a long distance phone call), and it’s clear that in times of war and for commercial espionage all other times, there have always been people who have made it their profession to break, or break into telecommunications networks. But those were different concerns which preoccupied the national telephone companies. The revenue lost by amateurs trying to obtain free long distance services could not have been more than an insignificant dent in their balance sheets. Commercial or national espionage would have been more of a concern to the companies or governments using the network, than to the operators themselves. Simply said, either the scale of the problem was small, or the reputation of the communications provider was not at risk. The term “Denial of Service” had not yet been invented.

That was then. A communications device was a phone, or a modem, or a teletype. The advent of the Internet changed all that. In the beginning the Internet and voice communications stayed separate for a while; the Internet remained for browsing, and file downloading, and e-mail, while the business of transmitting voice stayed in the Time Division Multiplex (TDM) networks. We soon learned what virus were and how Denial of Service (DoS) attacks could cripple a bank, or a shopping site, or the government site of a small Nordic country. But still, these problems could not cross into the TDM domain. There was too much of a difference in the technologies built a natural firewall. But it didn’t last long. What at first was just a curiosity, turned into a legitimate commercial interest for those in a garage who could understand the technology and put it out there to be used by anyone. Voice over IP (VoIP) was born.

In the span of 10 years, VoIP has torn down the natural firewall between the Internet and voice communications. As soon as some of the first start-ups became large enough to have a reputation and a commercial interest to protect, as soon as a small enterprise threw out their analog Private Branch Exchanges (PBX) for a server which could now take care of all the companies telephone needs, as soon as the large operators and cable companies became aware that they needed to face the new technology and adopt it themselves, the security threats of the internet became their new threats. Telecommunications security took on a whole new meaning and acquired an entire new vocabulary.

The author has worked in telecommunications for more than twenty years, for about the last 8 in Internet Protocol (IP) communications or what is commonly called *Next Generation Networks (NGN).* Even though security in these networks was not entirely neglected in the first few technologies, it was not the top-most priority either. In the last three to four years, the last group mentioned above, the large operators and cable companies started to take real interest in VoIP for commercial use, and not just to investigate it in a laboratory. It is these companies that are putting security at the top. At the same time, they are also learning about the new threats, the standards, the defense mechanisms, and the products. Some are basically learning as they go. It is clear that a methodology needs to be developed to assist operators with the task of securing a next generation network. That is what this thesis aims to study and develop.

1.2 Telecom Provider Security Concerns

Regardless of the technology, the greatest concern for a telecommunications company is the loss of the large amounts of revenue that under normal circumstances it takes for granted. This can come not just via theft of service, in fact, even though this will steal revenue from an operator. The size of the loss from theft of service would generally be insignificant compared to two other ways of not generating income: loss of customers, current and future, due to tarnished reputation, and loss of service due to massive attacks on the network control.

With the migration of telecommunications networks to an all-IP technology and architecture, operators are having to dedicate a lot more money and resources to the area of security. This has been reflected within telecom companies and cable operators in the following areas:

Education – it is highly probable that many of the employees of these companies, who are engaged in the maintenance, design, and operations of the network infrastructure left college when the first personal computer had not yet been invented. Even those that specialized in data communications as opposed to voice, have most likely dealt with technologies and standards long discarded. As we will see in this paper, the Internet, cellular communications, and voice over IP have spawned hundreds of standards, concepts, and new acronyms and terms. In order for the operators to deal successfully with the security challenges coming their way, re-training of large numbers of their technical employees has been mandatory.

Consulting services – a key staple of the procurement process in telecommunications, as indeed in almost any commercial undertaking, is the process of public tendering of contracts. This has been the case also with the previous telecommunications technologies. A tender usually begins with a document which is sent to participating companies called the Request for Quotation (RFP), or Request for Pricing (RFP), sometimes preceded by a Request for Information (RFI). The difference between the RFI and the previous two is that the latter does not require pricing information. With next generation networks, it is often seen that these documents are written by the operator with the assistance of consulting companies, who have acquired the talent in the diverse topics of NGN such as security and quality of service. Consultants will usually be under contract for the duration of the tendering process, while an operator's own employees learn the new technology.

Broader spectrum of products and technologies – along with new technology, standards, and acronyms come a number of products that would have had no place in a telecommunications network before. These products are sold by companies, some of which did not exist a few years ago, and some of which will not be around a few years in the future. The products are not only hardware and software but also consulting and integration services. Understanding this new “ecosystem” is vital for the security experts in such an organization.

Constant upkeep – finally, all this new technology has something in common, it never stops changing: new applications, new releases, new features, new platforms, as well as: new viruses, new attacks, new devices from which to mount the attacks, and new threats. In summary, security is a long term undertaking.

1.3 Thesis Outline

This work is divided into four parts:

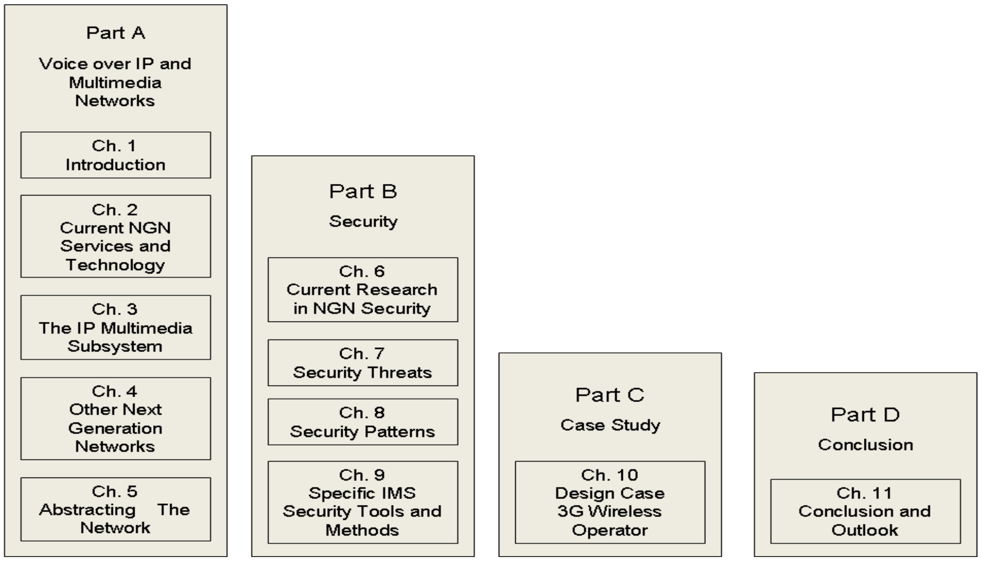

Part A
Voice over IP and Multimedia Networks
Ch. 1 Introduction
Ch. 2 Current NGN Services and Technology
Ch. 3 The IP Multimedia Subsystem
Ch. 4 Other Next Generation Networks
Ch. 5 Abstracting The Network
Part B
Security
Ch. 6 Current Research in NGN Security
Ch. 7 Security Threats
Ch. 8 Security Patterns
Ch. 9 Specific IMS Security Tools and Methods
Part C
Case Study
Ch. 10 Design Case 3G Wireless Operator
Part D
Conclusion
Ch. 11 Conclusion and Outlook

Figure 1-1 Thesis Outline

in part A we dedicate 5 chapters to Next Generation Networks and their properties. In Chapter 1, we’ve given the general motivation for this research and the main areas of concern for an operator regarding the circumstances around the problem of network security. In Chapter 2, we set the background for the work to follow by presenting the state of the telecommunications industry today. We start with the subscribers to communications services, the consumers, how their habits have changed, how the way of accessing the services have progressed, and what their new expectations are. We then continue with the providers of the communications services, the different and competing types of operators that are vying for consumers, and the specific difficulties that they each face as a result of the different technologies they use. We finally introduce the different standardization bodies advancing the technology, as well as take a first look at the dizzying array of recommendations regarding security.

In Chapter 3 we introduce the IP Multimedia Subsystem (IMS) architecture, looking at its basic components, interfaces, brief history, and design principles. We then provide examples of what operators are introducing IMS, proving that IMS has ceased to be just the latest new acronym in telecommunications in order to become a reality.

In order to have a frame of reference for IMS and the work to follow, we provide in Chapter 4 brief descriptions of other next generation network frameworks, developed in parallel to IMS, and even prior to it, which have now come to embrace and adopt many of the ideas behind IMS.

In Chapter 5 we introduce a new idea with respect to next generation networks, what we may call a *generic NGN*. We use this generic or abstract view of the NGN in order to separate those aspects of the components of a next generation network which have a bearing on the security risks they present. In Chapter 10, we will use the results of this analysis to help with our security design.

In Part B we introduce the topic of security in NGN’s, in 4 chapters. We start in Chapter 6 with a review of the work being done in this and related areas. Chapter 7 dissects an IMS network and analyzes the types of security threats that different domains and interfaces of the network will face. Chapter 8 introduces the topic of Security Patterns and presents 6 new patterns. In Chapter 9 we look at some existing security patterns, previously published in the literature, and we start to see how the complete set of patterns from chapters 8 and 9 might be used for the task at hand.

Part C is the culmination of this work. In a single chapter, Chapter 10, we introduce a fictitious, but quite representative operator, *Alpha Multimedia Telecommunications (AMT)*; we present its current network and plans for expansion into a Third Generation (3G) mobile architecture. We give examples of several new applications which AMT will offer via its new network. We then proceed to design a security architecture using the knowledge and tools gained and developed in the previous two parts.

We conclude in Part B with one chapter about conclusions and ideas for future work. An extensive list of references used finishes this paper.

2 CURRENT NGN SERVICES AND TECHNOLOGY

Security is not just an add-on to a network like just another application or service. Security in networked systems has the purpose of protecting assets and information. In order to see what type of assets and information it is that needs to be protected, it is helpful to review who the consumers and who the providers of that information are. In this chapter we review the state of the technology from the consumer point of view. We then take a look at the providers: the incumbent fixed line and cellular telephone companies (“telcos”), the cable companies, and then the new competitive operators. We then review the standards that the operators can use to help them provide new services, including security, and some legal mandates that are applicable to all communications operators. Finally, we summarize some of the relevant security related standards.

2.1 Consumers

There have never been more choices for communications technology and communications services providers than today. For voice communications there are fixed line phones, cellular phones, voice over IP over cable, voice over IP over Digital Subscriber Line (DSL), voice over IP over WiFi, voice over IP over cellular data services, Push-to-Talk, Vonage, Skype®, GoogleTalk® and others. For text communications there is Short Messaging Service (SMS), Instant Messaging (IM), e-mail from a fixed location, e-mail from the mobile phone, with multiple varieties and providers of each. For video services, including real time communications and video retrieval there is video calling (albeit with limited availability), video streaming, web video conferencing, as well as peer-to-peer video services like Skype®, “see-what-I-see” one way video, YouTube® and other group sharing communities. With broadband cellular such as 3GPP’s Long Term Evolution (LTE) and other wireless technologies like WiMax becoming more prevalent in the next coming years, the choices will increase even more, and a user may not even know, let alone care about, via what facilities the communication takes place.

All this availability of electronic communications for leisure, for business, and for commerce is causing shifts in usage and habits by consumers. Unpublished internal studies by a major cell phone manufacturer indicate that the average 25 to 35 year-old spent an average of 48 minutes a day using his or her mobile device in 2007, versus 30 minutes a day in 2006, a clearly increasing trend. The top four applications were messaging, multimedia, browsing, and voice calls in that order. In 2005, only 7.7% of consumers had given up their land-line for wireless only service. By 2007 it was 15.8%. [CTI08] The number of SMS messages exchanged by users in North America per month in 2005 was 7.2 billion; in 2007 it had gone up to 75 billion [CTI08]. One might think that users excited by all this choice and availability of new services might not give much thought to how secure their use is, but this is not the case. A study by Harris Interactive in September 2008 [HAR08] has found that even among teens, 53% of those surveyed reported that the "security that guarantees only you have access to data on your phone" is "very important" or "absolutely essential".

Regardless of the service available at one's fingertips, the service provider one uses, or the type of access network through which all these services can be funneled, the volume of data and the types of applications sending and receiving the data will continue to increase, as well as the diversity of sources and locations where the applications reside. No longer will the local phone company have every piece of the puzzle under its control. The more the networks grow and the more these services are taken for granted, the greater need for robust security to protect the user, the data, the network, and the application servers from attack.

2.2 Operators

2.2.1 Telcos and Wireless

In the first days of cellular service, the same telephone company used to provide both wireline and wireless services. Later, in the height of the dotcom boon, most telephone companies divested themselves of the wireless operations, and became separate companies. The trend did reverse itself somewhat later, but we're still in an environment where there is a separation between networks for wireless service and networks for wireline service, in operations, provisioning, and security. Another legacy, but this time stemming from the divestiture of the AT&T telephone monopoly, is the separation of long distance networks and local Class 5 operations.

In this environment of fractured networks, separation of operations, and also of increasing competition from cable and Competitive Local Exchange Companies (CLEC's), traditional telephone companies and wireless providers are trying to maintain their customer base, and add features and differentiators to their services to make them "sticky".

- Telephone companies are trying to expand into video into the home by deploying billions of dollars worth of fiber [MAY06].
- They have over years deployed DSL in higher and higher bandwidths in order to provide broadband internet service.
- They have allied with satellite TV providers in offering bundled packages to give the "illusion" of a single source of services.
- Some cellular providers have begun to offer also VoIP to the fixed line, competing with Vonage and the cable companies [TAY08].
- Some cellular providers have started to offer dual mode services, where a two-radio cellular phone, with GSM and WiFi capabilities can make/receive calls either via the GSM access network, or when in the home, use the wireless WiFi home network and VoIP [BAR06].

- They have developed distinctive wireless plans that offer certain benefits (e.g. T-Mobile’s myFaves®).
- They are starting to provide multimedia features over cellular broadband, including video, gaming, and soon Rich Communications Suite (RCS), an as yet unpublished standard by a the RCS Initiative group of companies [NOK08].

Telephone companies and cellular providers want to be able to deploy all these features, and more to come, in a way which does not duplicate basic pillars of each service, like billing, security, quality of service, provisioning, and maintenance.

Telephone companies also would like to control what is flowing over subscribers DSL lines and their wireless networks (witness the blocking by T-Mobile U.K. of VoIP and IP Messaging [TMC06]), like some cable companies have tried to do, as in this case by Comcast in North America in 2007 where peer to peer traffic was monitored and limited [SVE07]. At the very least, they would like to have the technical capability to do it; whether this is legal or not is for the courts to decide. One reason for this need is that the owners of the local loop may want to offer services which compete with the likes of Skype, Vonage, etc. The telephone company might want to provide better QoS to its own VoIP subscribers than those of the competing provider.

On the other hand telephone companies, both wireline and cellular, have very little incentive to mothball their still working digital switches, which still work well, have been fully depreciated, and are known quantities in terms of security, and other measures of quality. So some of them are trying to delay the inevitable technology change, by gradually evolving their existing networks without radically switching to the next generation technology.

Some options for adding new multimedia features and keeping at least part of the existing infrastructure are described below:

Public Switched Telephone Network (PSTN) Emulation – The European Telecommunications Standards Institute Telecommunications and Internet converged Services and Protocols for Advanced Networking (ETSI TISPAN) has defined a subsystem, as part of its overall NGN architecture, which enables the existing millions of deployed analog telephones, fax machines, and telephone systems (PBX's) to still receive services from an all-IP next generation network like IMS. It is recognized by the organizations that make up ETSI, that all these devices will not go away for a long time to come, and that not all users are "early adopters". The purpose of the PSTN Emulation Subsystem (PES), as it's called, is then to emulate the PSTN features and "look and feel" of those features such that the user (and the terminal) does not know that it's being served by an NGN. In a trade publication [HIL05] it's explained as follows: "the idea of PSTN Emulation is to create a service in an NGN that is effectively identical to the PSTN, with the same feature set and user ergonomics. This means that, as far as the end user is concerned, nothing has changed." PSTN Emulation is defined in [ETS182].

PSTN Simulation – In PSTN Simulation, the operator offers the users a service which is not completely equal to the legacy PSTN services, but is not 100% new either. In the same trade publication as above [09] it's described as follows: "PSTN Simulation provides something that looks generally like a PSTN or Integrated Services Digital Network (ISDN) service, but doesn't resemble it in all respects. For example, it can use a variety of new terminal types, offer new value-added features, but also not offer some old ones. Simulation is more about allowing evolution into a new NGN environment than replicating the old environment exactly". PSTN Simulation services are defined in a number of ETSI TISPAN documents [ETS00]

There are clearly a number of reasons why an operator might choose to go with one or the other strategy, when migrating to a next generation network. A discussion of the advantages and disadvantages can be found in [HIL05]. That is not relevant to this work. All we are interested in here are the repercussions that those decision may have on the security infrastructure to be designed.

Generic Access Network (GAN) – The GAN is an alternative for wireless Global Switching Mobile (GSM) operators which want to enhance the regular cellular service of their voice subscribers, by allowing them to make and receive calls also via a home wireless WiFi network, or via a WiFi hotspot [BAR06]. This can be achieved by using a dual mode telephone, i.e. with 2 radios. This technology has benefits for both the operator and the user. For the operator, using a WiFi network or hotspot unloads the cellular access network, carrying traffic via an IP backbone instead. For the user, it has the benefit of better quality of voice in areas where the cellular radio signal is very weak, as it often happens in some out-of-the-way residential areas. It also may mean, depending on the pricing plan, that no airtime minutes are used when the call is over the WiFi. The design also allows for seamless handover of an active call from one network to the other.

In a GAN, which is commercially referred to as the Unlicensed Mobile Access (UMA), and is defined in 3GPP spec TS 43.318 [3GP318], the Mobile Station (cell phone) tunnels the GSM signaling protocols and voice and data traffic over an IP connection to a network element, the Generic Access Network Controller, which looks to the core network as just another Base Station, which also includes the necessary intelligence to do mutual authentication, encryption and data integrity for signaling, voice and data traffic.

Mobile Access Gateways – These network elements can also be called Wireless Access Gateways. No definition of them has yet been inserted in the 3GPP standards but some infrastructure vendors already claim support in their products. This type of gateway allows a legacy 2G/3G radio access network to connect directly to a next generation network (e.g. an IMS), bypassing the traditional Mobile Switching Centers of the cellular network. It does this by converting the signaling from legacy protocols using Signaling System Number 7 (SS7) into the Session Initiation Protocol (SIP) [ROS02], and by converting the media from PCM TDM (Pulse Code Modulation Time Division Multiplex) into an IP media stream encoded with one of the standard International Telephone Union’s (ITU-T) codecs (G.711, G.721, G.729, etc.). It should be observed that the use of this access gateway benefits in reality only the operator. The user sees no difference (or should not see any) since he or she is still using the same 2G/3G mobile handset and has access to no more applications than before. For the network operator the benefit is one of being able to start to move in the direction of all-IP network core for control and for media, while still being able to keep the same subscriber base using the same devices, i.e. no disruption of services. An operator might want to move in this direction if it has a business need to expand capacity and does not want to invest into additional TDM equipment, soon to be outdated. An operator might also want to add new services in an overlay fashion, in other words, continue with the traditional voice services for a segment of the subscriber base, but add other applications piecemeal for those “early adopter” subscribers who want to upgrade their devices to all-IP.

The security aspect of Mobile Access Gateways is related to the transport of media and signaling between the gateway and the IMS core. Additionally, authentication of subscribers may be done in the gateway, depending on which way the standards go.

2.2.2 Cable Companies

Out of more than 112 million TV households in the United States today, Cable companies serve about 68 million or 58% [NCT08]. Out of those, 50 million are premium cable units, meaning they are equipped to receive digital broadcasts and broadband data. 37 million of these households subscribe to high speed internet, and about 16 million subscribe to VoIP telephony [NCT08]. With a total of 117 million homes “passed” by (i.e. capable of subscribing to) cable high-speed data service, the overall population potential for data and VoIP subscribers is still large for most cable companies.

Cable companies expanded into offering telephony and internet services from a position of strength in video, their initial product. Some of them started out with traditional digital switching with large software controlled central offices, like the telephone companies. Later, some moved into softswitching by capping their legacy switches and deploying server based IP telephony. Now once again, some of them are migrating to an IMS architecture, since CableLabs, the research consortium funded by the North American cable companies, has adopted the SIP based architecture as the foundation of the PacketCable 2.0 standard.

For cable companies, voice is therefore the third main service on their menu and cellular communications will be their fourth.

Cable companies have networks which are substantially different from those of the telephone companies, consisting of an IP backbone connecting independent fiber optic rings, from which coaxial cable provides the access to the home. In the past years, the cable companies have invested vast amounts of money to improve their networks, mainly the coaxial segment, in order to allow it to carry greater bandwidths of data needed for High Definition Television (HDTV) and broadband internet [STO03]. That improvement is now largely complete.

Due in part to the different network architecture, the cable companies have a competitive advantage in moving to an IMS based voice and multimedia services architecture: the hybrid fiber coax network is in a way not very different from an Ethernet based IP backbone, and cable companies do not have hundreds of legacy features implemented over 30 years, which their customers (albeit not many) may still expect to have access to in an IP-based network.

Since the IMS is also the core network for multimedia services, the advantage that cable companies also have with regards to content (movie libraries) is an additional incentive to develop new and differentiated applications.

Lastly, cable companies are interested also in offering wireless services by executing roaming agreements with cellular providers, while they deploy their own 3G or 4G radio access networks. With fixed mobile convergence, most wireless calls (those in and around the home) would not use the wireless network but rather the cable companies’ own IP backbone.

2.2.3 New entrants

Telephone companies and cable companies are not alone in providing new IP based multimedia services including VoIP. In fact the first providers of these services were not the established operators but rather new players such as Vonage, Skype, Lingo, Voip.com, Google, etc. For some of these, VoIP is the reason they exist. Others just have VoIP as an added bonus to their existing menu of applications, and voice itself is limited in features, reach, and quality of service guarantees. Also restrictions like limited 911 service and the lack of name recognition or standing in the eyes of some consumers, will mean that these providers will not likely get to be a real threat to cable or telephone companies. As previously mentioned, the fact that these new entrants provide their service over facilities which they do not control (the telephone or cable companies copper or coax plant) will mean that they cannot control some aspects of the service such as quality of service or parts of the end-to-end security.

2.3 Standards

2.3.1 3GPP/3GPP2

The Third Generation Partnership Project (3GPP) was formed in December 1988. Its initial purpose as defined in the initial agreement documents [3GP08] was to "co-operate for the production of a complete set of globally applicable Technical Specifications for a 3rd Generation Mobile System based on the evolved GSM core networks and the radio access technologies supported by 3GPP partners (i.e., UMTS Terrestrial Radio Access Network (UTRAN) both Frequency Domain Division (FDD) and Time Domain Division (TDD) modes)". The "partners" in 3GPP are Organizational Partners, defined as "a standards organization with a national, regional or other officially recognized status in their country or region" and Market Representation Partners, defined as "organizations invited to participate by the Organizational Partners to offer market advice to 3GPP and to bring into 3GPP a consensus view of market requirements (e.g. services, features and functionality) falling within the 3GPP scope" [3GP07].

Initially, 3GPP defined four Technical Specification Groups, (TSG), to work in four distinct areas: Radio Access Network, Core Network, Terminals, Service and System Aspects. Many aspects of security fall under the last one, although there are some areas, like authentication, which are also addressed in other groups.

One of the goals of 3GPP in designing a replacement architecture for the core network, which until now has been heavily circuit switch oriented, was to make as much use as possible of existing Internet standards and protocols, developed and maintained by the Internet Engineering Task Force (IETF). It was this goal, that lead the Services and System Aspects TSG to choose the Session Initiation Protocol early on as the session control protocol for the IP Multimedia Subsystem. The SIP protocol is defined in IETF Request For Comments (RFC) 3261 [ROS02]. It is a text based protocol which borrows many characteristics from the Hyper Text Transfer Protocol (HTTP). As the name implies, SIP is used to initiate, modify, and terminate sessions among two or more participants. Session information (end point capabilities, end point address, type of session, etc.) is exchanged by using the Session Description Protocol (SDP), embedded in SIP messages. SDP is defined in RFC 2327 [HAN98]. 3GPP has defined a number of additional RFC's in addition to RFC 3261 as a requirement for IMS networks.

The IMS, which is only one of the areas being developed by 3GPP, was introduced in 3GPP Release 5. Release 7 specifications have been frozen since the second half of 2007. Release 8 is currently on-going.

A good summary of the history of the IMS in 3GPP and of 3GPP/IETF collaboration can be found in [CAM08].

2.3.2 TISPAN

The Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN) standardization body was created by the European Telecommunications Standards Institute (ETSI) in 2003 to contribute to the standardization of Next Generation Telecommunications networks

[TIS00]. Whereas 3GPP had its origins in the cellular world, TISPAN had its roots in fixed networks and its creation was due in part to the need for harmonizing the fixed telephony and Internet architectures.

TISPAN adopted the 3GPP's IMS architecture, based on SIP, in its NGN Release 1 (December 2005) and added to it the necessary functional blocks to adapt it to fixed, legacy networks. In early 2008 two important activities were finished: the first one was the completion of NGN Release 2 and the second was the transfer to 3GPP of all the common IMS specifications, in order to have only one standards body, 3GPP, responsible for the IMS core.

Within TISPAN, working group 7 (WG7), is responsible for security aspects. The responsibilities of WG7 are [TIS08]:

- Conducting studies leading to deliverables on security;
- Management and co-ordination of the development of security specifications for the next generation telephony and multimedia communications;
- Investigation of security services and mechanisms required for providing services over the Internet;
- Development of security analyses of candidate protocols and network elements to be used within the NGN framework to implement capabilities;
- Tracking ongoing worldwide security activities of interest to TISPAN

2.3.3 PacketCable

PacketCable™ is one of 8 projects currently in existence within CableLabs® [CAB00], the consortium founded in 1988 by the cable operating companies, to research, develop, and test new cable telecommunications technologies. PacketCable’s goal is to “develop interoperable interface specifications for delivering advanced, real-time multimedia services over two-way cable plant” [PAC00].

PacketCable has been through 4 releases since its beginning: PacketCable 1.0, 1.5, PacketCable Multimedia, and PacketCable 2.0.

PacketCable 1.0 and 1.5 define an infrastructure for delivering packet based residential telephony services over the data channels in existing coaxial cable to the home**.** This end-to-end architecture covers all aspects of providing basic telephony service, including provisioning, call signaling, call detail recording, configuration management, quality of service, interconnection to the PSTN, and security [PAC01]. As opposed to other VoIP services, the system that PacketCable 1.0 and 1.5 define is for phone-to-phone service, using a type of media gateway control protocol, the Network Control Signaling, or NCS, to control residential media gateways. These gateways are the interface between a typical “black” phone using an a regular network connector cable for analog signaling, and the cable network using coaxial physical transport and packet data transmission. PacketCable 1.0 and 1.5 do not provide for user mobility or multimedia capabilities, but has been tremendously successful in North America, with 16 million VoIP lines deployed by the end of 2008.

PacketCable Multimedia (PCMM) was an effort to introduce a generic quality of service infrastructure for any type of applications, including voice and multimedia. PCMM introduced the Policy Decision Function (PDF) and the protocols between it and other network elements involved in specifying or delivering quality of service guarantees. The PCMM architecture has not been widely deployed within the cable industry.

PacketCable 2.0 is the current standard being developed by CableLabs and deployed into the network by the cable companies. PacketCable 2.0 embraces fully the concept of 3GPP IMS and adopts its key network components, protocols, and methods such as for Security, Quality of Service, and Billing, although these are tailored towards the needs of cable networks. In fact, CableLabs is a member of the 3GPP consortium and its requirements are being worked into the relevant core technical specifications. Although no commercial deployments exist as of this writing which use PacketCable 2.0, the first trials are in progress and some of them will most likely turn commercial in 2009.

2.3.4 IETF

The Internet Engineering Task Force was started more than 22 years ago in San Diego, California, as a “a loosely self-organized group of people who contribute to the engineering and evolution of Internet technologies” [HOF06]. It became part of the Internet Society (ISOC) in 1992. According to [HOF06] its mission includes:

- Identifying, and proposing solutions to, pressing operational and technical problems in the Internet
- Specifying the development or usage of protocols and the near-term architecture to solve such technical problems for the Internet
- Making recommendations to the Internet Engineering Steering Group (IESG) regarding the standardization of protocols and protocol usage in the Internet
- Facilitating technology transfer from the Internet Research Task Force (IRTF) to the wider Internet community
- Providing a forum for the exchange of information within the Internet community between vendors, users, researchers, agency contractors, and network managers

Although there is no formal membership to the IETF, ISOC has more than 28,000 individual and over 80 organizational members. The IETF is organized into 8 Areas, each area containing anywhere from 1 to almost 30 different Working Groups. The Security Area contains 17 Working Groups; some of the most relevant to our work here are: *ipsecme,* dealing with IPSec maintenance and extensions, *tls,* dealing with transport layer security, *keyprov,* which looks into symmetric key provisioning, and *krb-wg,* in charge of Kerberos issues. The other areas are: Applications, General, Internet, Operations and Management, Real-time Applications and Infrastructure, Routing, and Transport.

2.4 Regulatory Mandates

Two features in telecommunications networks are important enough to be mandated by the government for deployment by service providers: Lawful Interception and Emergency Services. Since they both impact the design of an overall next generation network security architecture, the current status and related standards are presented here briefly

2.4.1 Lawful Interception (LI)

Telecommunications networks of all types are subject to governmental regulations regarding the legal interception of communications. Next Generation Networks, including IMS, are no exception. Legal interception requirements impose certain architectural enhancements, new network elements, and new protocols for the collection of session information which, for both signaling and media, like the rest of the network, must meet the same strict security criteria as regular, not intercepted sessions. Another key requirement of LI architecture is that there cannot be any external indication (i.e. to the person or persons whose session is being intercepted) that electronic surveillance is taking place.

In North America, there exist several standards which dictate how operators, and therefore also vendors of telecom equipment, must provide for the legal interception of all types of sessions. Below we list those that are applicable to the networks described in this paper.

ATIS 0700005-2007 – This Alliance for Telecommunications Industry Solutions (ATIS) standard specifies the requirements for UMTS (Universal Mobile Telecommunications Services) VoIP sessions, which would also apply to any network using an IMS core.

ANSI J-STD-025-B-2006 – This Telecommunications Industry Association (TIA) and Alliance for Telecommunications Industry Solutions (ATIS) joint standard specifies the requirements for wireline and wireless communications.

T1.678 (ATIS-1000678.2006) – This ATIS standard provides the requirements for interception of VoIP in wireline telecommunications networks. It includes support for supplementary services such as multi-party calls, call transfer, etc.

PKT-SP-ESP-I03-40113 – This is a CableLabs standard covering Lawful Intercept for PacketCable 1.1 VoIP networks.

PKT-SP-ESP1.5-I02-070412 - This is a CableLabs standard covering Lawful Intercept for PacketCable 1.5 VoIP networks.

PKT-SP-ES-DCI-I01 and PKT-SP-ES-INF-I02 – These are CableLabs standards covering Lawful Intercept for PacketCable 2.0 VoIP network.

TIA-1066 – This TIA standard covers the requirements for CDMA2000 VoIP networks.

WTSC T1.724 Rel. 5 - UMTS and TIA-1072 – These standards by the ATIS Wireless Telecom Systems Committee (WTSC) and by the TIA, respectively, address the requirements for Push-to-Talk over UMTS/GPRS and CDMA2000, respectively.

All IMS networks being deployed in North America must meet the relevant standards listed above from day one. Although standards vary, essentially this means the capability of intercepting session signaling (IRI - Intercept Related Information) and media (CC – Call Content) and delivering it securely, in real time, to one or more law enforcement agencies. All above standards have requirements regarding the handling, keeping, safeguarding, and destroying of the information, in order to prevent its unauthorized use. Due to the complexity of the requirements, both technical and regulatory (court order processing, competing law enforcement receiving agencies, etc.), some operators elect to outsource lawful intercept functions to specialized companies. They still must provide the interfaces to their network however, via which the companies can activate the intercepts and retrieve the data. These standard interfaces are known as the "H" (handover) interfaces and include an interface for Administration of the intercept instructions (H1), for sending the IRI to the enforcement agency (H2), and for sending the CC (H3).

The security designs developed later in this paper must take into account the Lawful Interception infrastructure put in place by the operator.

2.4.2 Emergency Services

In 1968, AT&T announced that it would establish the digits 9-1-1 (nine-one-one) as the single code for emergency calls throughout the United States. Ever since, 911 services has been a critical component of telecommunications networks. Every year approximately 240 million 911 calls are made in the United States. [NEN00].

The current 911 system however, was designed to enable circuit switched calls to emergency services, not data, and the present architecture is in danger of not being able to work with new technologies. Today, usage patterns by subscribers are changing: it is estimated that between 23 and 37 % of US wireless subscribers will use their cell phone as their primary communications device by 2009 [CTI09]. Trends also suggest that there will be more than 27 million residential VoIP subscribers by 2009 [NCT08]. Cities like San Francisco and Philadelphia are in the process of deploying citywide WiMax networks, which will enable users with WiFi calling plans to dial emergency services from anywhere in the city, requiring new methods of pinpointing their location.

The National Emergency Number Association (NENA) has been planning for the modernization of the national Emergency Services standards needed to confront these changes in technology, mobility, and usage patterns. In 2005 NENA put out the Interim VoIP Architecture for Enhanced 9-1-1 Services, normally referred to as i2 [NEN05], in order to “develop the architecture to support the interconnection of VoIP domains with the existing Emergency Services Network infrastructure in support of the migration toward end-to-end emergency calling over the VoIP networks”. This document was complemented in 2007 with the publication of “NENA Functional and Interface Standards for Next Generation 9-1-1 Version 1.0 NENA 08-002 Version 1.0” [NEN07] commonly referred to as i3.

Both of these standards include requirements in terms of the Security architecture, which will need to be taken into account when designing an overall security infrastructure later in this work.

2.5 Current Security Standards

Security in today’s IP telecommunications and multimedia networks is a vast area, from authenticating a subscriber which no longer has a direct one to one relationship with a physical line, to protecting hundreds of heterogeneous applications in so many servers from attacks which keep mutating. The size of the problem means that many of the security designs around a new green-field or an evolving network, are not complete before launch, and keep evolving in an ad-hoc way. Originally, the initial security architecture is put into place, and as new threats are recognized, or as a reaction to detected attacks, successive components are added.

Operators do attempt to gather enough information about the manufacturers’ capabilities with respect to security. Requests for Proposal (RFP’s) with security sections with questions numbering in the hundreds are not uncommon. Operators, or their consultants, gather all the possible standards documents which they think are applicable to the network they plan to build and question the bidders point by point about the requirements therein. The problem is that usually these documents and requirements are targeted at protecting a very specific asset or protocol or interface, and do not (cannot), take the entire network into account. But not only are there documents dedicated specifically to security; often, a given standard will have its own security section embedded, meaning that the recommendations for protecting the network will not be all gathered in one document but scattered over many.

The following are examples of security guidelines and documents for next generation IP networks (there are too many to list them all). The purpose of this list is not to describe or summarize what each of these documents specify, but just to give the reader an idea of the scale of the task before the operator who wants to secure the network according to the existing standards. The idea here is that it would be the wrong approach to go through every applicable standard and decide how to implement it. This paper proposes a different methodology which will be evident in subsequent chapters. The areas covered below would be the standards applicable to mobile, cable and fixed, operators respectively.

2.5.1 3GPP

The following are the most important security specifications defined by the 3GPP:

TS 21.133 - 3G security; Security threats and requirements (only up to 3GPP Release 4, after that, security info contained in individual specs).

TS 29.204 - Signalling System No. 7 (SS7) security gateway; Architecture, functional description and protocol details.

TS 29.800 - Signalling System No. 7 (SS7) Security Gateway; Architecture, functional description and protocol details.

TS 32.371 - Telecommunication management; Security Management concept and requirements.

TS 32.372 – Telecommunication management; Security services for Integration Reference Point (IRP): Information Service (IS).

TS 32.373 - Telecommunication management; Security services for Integration Reference Point (IRP): Common Object Request Broker Architecture (CORBA) Solution.

TS 32.375 - Telecommunication management; Security services for Integration Reference Point (IRP): File integrity solution.

TS 33.XXX – The “33” series documents concern security. There are 44 of them as of this writing. Some examples are: Cryptographic algorithm requirements, Lawful interception requirements, Generic Authentication Architecture (GAA), Network Domain Security (NDS), Liberty Alliance and 3GPP security interworking, and many more.

2.5.2 CableLabs PacketCable

The following are the security specifications defined by the PacketCable and Data Over Cable Service Interface Specification (DOCSIS) projects of CableLabs:

PKT-SP-33.203-I04-080425 - PacketCable™ IMS Delta Specifications 3G security; Access security for IP-based services Specification 3GPP TS 33.203.

PKT-TR-SEC-V05-080425 - PacketCable™ Security Technical Report.

SP-SECv3.0 - Security Specification

2.5.3 ETSI TISPAN

The following are some of the security specifications defined by TISPAN:

TR 185-008 - Analysis of security mechanisms for customer networks connected to TISPAN NGN R2.

TR 102-419 - Security analysis of IPv6 application in telecommunications standards.

TR 202-549 - Design Guide; Application of security countermeasures to service capabilities.

TR 202-387 - Security Design Guide.

TR 102-165-1 - Methods and protocols; Part 1: Method and pro-forma for Threat, Risk, Vulnerability Analysis.

TR 102-165-2 - Methods and protocols; Part 2: Protocol Framework Definition; Security Counter Measures.

ES 202-382 – Security Design Guide; Method and pro-forma for defining Protection Profiles.

ES 202-383 – Security Design Guide; Method and pro-forma for defining Security Targets.

In this chapter we have set the scenario by briefly considering the consumer point of view; expectations, types of applications that will be available to users, including the devices that will be used for those applications, which will in many ways dictate the security requirements. We have also seen where the operators are coming from and where they are going, and what this means as far as their experience with the type of security threats they will see in an NGN. Finally, we looked at the technologies and standards available to operators, including the vast literature with which they need to be familiar, in order to know which defenses should be deployed, and how.

In the next chapter will describe IMS in some detail, the framework in question being used in the case study to be analyzed. We will also give examples of some actual deployments or deployment announcements by operators world-wide.

3 THE IP MULTIMEDIA SUBSYSTEM (IMS)

The IP Multimedia Subsystem (IMS) is an architecture which defines the functional units, the interfaces, and the procedures for an all IP communications network. By communications it is meant voice, text, and multimedia sessions. The IMS was defined by the 3rd Generation Partnership Project consortium (3GPP), a group of standards organizations and other members, as the target architecture to replace existing circuit-switched centric networks, originally only those networks serving GSM cellular users, but more and more, others as well.

3.1 IMS Basics

According to [3GP228], the IMS, also referred to as the IP Multimedia Core Network (IP CN), enables cellular operators to "offer their subscribers multimedia services based on and built upon Internet applications, services and protocols". It attempts to bring the internet and mobile communications together and allow mobile users to benefit from the growth in the internet and applications provided therein. For that purpose, 3GPP has adopted where possible internet (IETF) protocols and standards.

In part by design, in part as a result of its adoption by other standards organizations as the accepted model for IP communications networks, the IMS Core Network can be accessed not only from the cellular providers' packet access networks (2.5G and 3G), but also from Wireless LAN access networks, fixed telephony networks based on IP, cable networks with data access capability, and corporate IP networks. The IMS is therefore said to be access agnostic.

The IMS does not define applications or services to be provided by operators, leaving those to the operators and third party application services providers. The IMS only defines session control functions, procedures, and reference points (i.e. interfaces) among them.

IMS has been in development since 2001, and is finally now reaching a high enough level of maturity for many operators to be considering it as the architecture of choice to replace existing deployed infrastructure [McG08]. The types of operators deploying IMS range from established fixed telephony operators looking to replace aging TDM networks, to wireless operators looking to take advantage of new third generation (3G) cellular wireless radio access networks, to new entrants looking to deploy directly 4G wireless access networks (whether LTE, W-CDMA, or WiMax), or combined satellite-terrestrial networks

IMS specifies a three layer architecture where transport, control, and services are separated by clearly defined interfaces or “reference points”. In addition, IMS specifies the different functions in the network assigning them to one of the three layers, and connects them to each other by means of the reference points. An architecture diagram representing the basic notions described above is given in Figure 1.

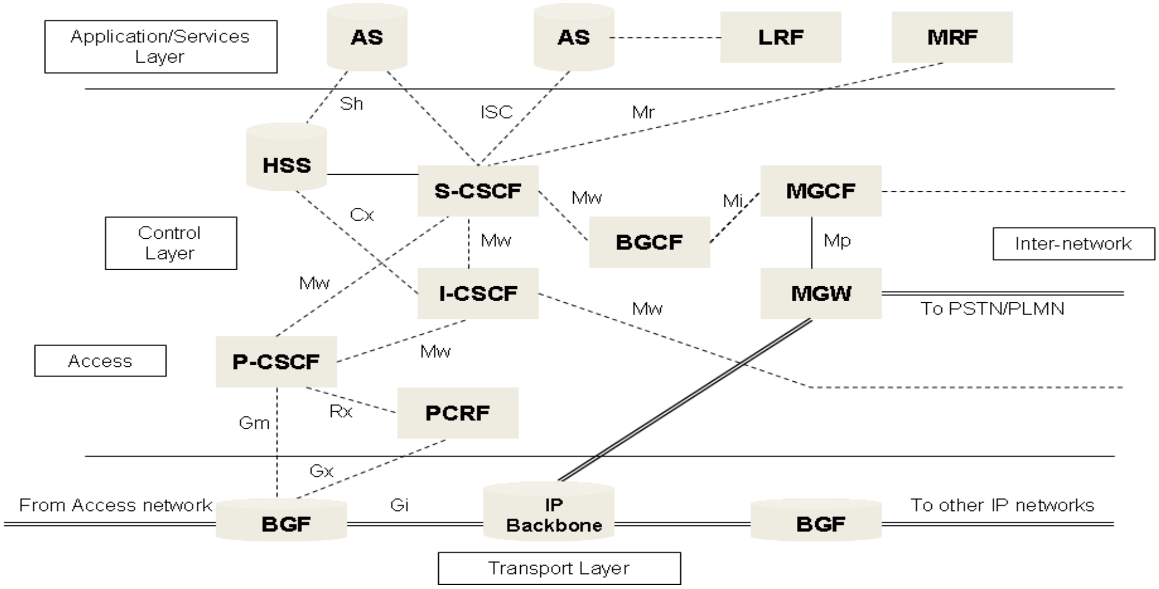


Figure 3-1 - Basic IMS Architecture

The diagram shows the basic logical functions of the IMS as specified in its main architecture standard document [3GPP002] as well as its most important signaling reference points (dashed lines), which are briefly explained below. Media follows the solid lines in the transport layer and to the Media Gateway (MGW). In order to make the following brief description of an IMS call more understandable, we list below for convenience the elements used, and their function:

Proxy Call/Session Control Server (P-CSCF): responsible for access side authentication and quality of service.

Interrogating Call/Session Control Server (I-CSCF): responsible for discovering a S-CSCF to assign to a particular subscriber.

Serving Call/Session Control Server (S-CSCF): responsible for registration and routing of call or session requests to application servers and to other networks.

Home Subscriber Server (HSS): database where all subscriber profiles reside

Breakout Gateway Control Function (BGCF): responsible for routing a call to the PSTN, via one of a number of MGCF's.

Media Gateway Control Function (MGCF): responsible for signaling with the PSTN using legacy protocols, and for controlling the MGW, which will connect the media.

Media Gateway (MGW): responsible for physically switching the media stream from IP to TDM and vice-versa.

We now give a brief introduction to how an IMS core network functions. An IMS subscriber with the right device (an IMS-compliant mobile or fixed terminal with a SIP client) registers with the network by sending a SIP REGISTER message to the Proxy CSCF (P-CSCF). The P-CSCF extracts the relevant data from the message and forwards it to the Interrogating CSCF (I-CSCF). The function of the I-CSCF is to contact the Home Subscriber Server (HSS) to find out, based on the properties of the subscriber and services subscribed, which Serving CSCF (S-CSCF) should be assigned to him or her. The I-CSCF then passes the message on to the right S-CSCF. The S-CSCF then proceeds to authenticate the subscriber according to its capabilities (more on this later) and to download the subscriber's profile from the HSS (i.e. what services is the subscriber allowed to use). Now the subscriber is registered with the network. The network knows where (to what device) to deliver incoming calls to him or her, and the subscriber can initiate calls or access applications.

An outgoing call to a user in the traditional telephone network (PSTN/PLMN) would proceed as follows. The user chooses the phone number from the phone book in the device and initiates a call. This causes the device to send a SIP INVITE message to the P-CSCF. The P-CSCF sends it on to the S-CSCF, which invokes the appropriate application server (AS) by looking at what kind of service is requested in the INVITE. The application server, in this case a voice application server, applies originating treatment (what telephony features need to be activated), and returns the message to the S-CSCF. The S-CSCF can then forward the message on to the Breakout Gateway Control Function (BGCF), which chooses a Media Gateway Control Function (MGCF) to interwork with the PSTN. In the message to the MGCF, an IP address and port for the calling device are specified, where the media path is to be connected. The MGCF instructs the Media Gateway (MGW) to connect that IP address and port via its internal matrix to a given port in its TDM side. From here on, the call is handed over to the PSTN.

There are other network elements which come into play in most sessions, and which are described below, but for an essential understanding of how a call is established via an IMS network, the above description will suffice.

3.2 Logical Functions

As the main contribution to the framework, understanding the different logical functions in the IMS is essential to grasping the overall architecture. There is a core group of functional units which make up the IMS and they are described next.

Home Subscriber Server (HSS) – The HSS is the main data store for subscriber information in the IMS, and for information which helps other elements of the IMS (e.g. the CSCF explained next) process incoming requests for services. Examples of subscriber information stored in the HSS are: unique private identifiers of the subscriber (either one or several), which are typically assigned to the end user device (e.g. SIP soft client, fixed SIP phone, etc.); unique public identifiers under which the subscriber can register with the IMS (one or several) which are typically associated with the "person", e.g. *john.smith_id_for_friends@ims.operator.com* or *john.smith_id_for_business@ims.operator.com.;* type of authentication to use with this subscriber; and services to which the user is allowed access. Any network element satisfying the interface requirements to the HSS, which will be described below, can retrieve or deposit (some of) the information stored in the HSS.

Call/Session Control Function (CSCF) – The CSCF is the Network Element (NE) responsible for session establishment, modification, and tear down. It uses SIP as its interface with the User Agent (UA) in the mobile terminal, and with other elements of the IMS network. The CSCF uses information downloaded from the HSS at mobile terminal registration time to route sessions (voice, text, media) to other UA's, to applications at the application layer, or to other IMS networks. The CSCF is typically a SIP proxy, as opposed to a softswitch implementing a Back-to-Back User Agent (B2BUA), meaning it acts on the SIP messages themselves instead of converting them to an internal protocol. This is however not mandated by the 3GPP. The CSCF can be of 3 different types: Proxy CSCF (P-CSCF), Interrogating CSCF (I-CSCF), and Serving CSCF (S-CSCF), depending on its position and function in the network. It's only important to note here, in order to understand the security mechanisms that will follow, that the P-CSCF is the one that faces (directly interfaces with) the access network, and that the I-CSCF is the one that faces the peer networks. Another important characteristic is that no media traverses the CSCF, or any other network core control element for that matter, only signaling.

Breakout Gateway Control Function (BGCF) – The BGCF selects the route an IMS call will take when the destination is in the circuit switched network. It is responsible for selecting both the network and the correct MGCF towards that network. The network can be the operator's own or that of a peer network.

Media Gateway Control Function (MGCF)/Media Gateway (MGW) – The MGCF is the network element responsible interfacing between the IMS realm which uses SIP, and the PSTN (Public Switched Telephone Network) or PLMN (Public Land Mobile Network), which use SS7 signaling. Sessions leaving the IMS domain towards the two legacy networks mentioned, or entering from there, are mediated by the MGCF. The MGCF, through which only signaling traverses, is also responsible for controlling the establishment of the physical media paths in the MGW. The MGW is mainly responsible for protocol conversion at layers 2 and 3 from Ethernet and IP on the IMS side, to TDM (Time Division Multiplex) in a T1 frame (24 64kb channels) or E1 frame (32 64kb channels) on the legacy side. It also needs to do trans-coding between the different codecs used. The details of these operations are not relevant but can be found in any introductory book on telecommunications. The MGCF and the MGW interface via ITU-T recommendation H.248.

Policy and Charging Rules Function (PCRF) – The PCRF is the element in the IMS responsible for policy and charging control. It interfaces with a policy and charging enforcement function (PCEF) in a router or border gateway, the element which the media physically traverses, and with the (P-) CSCF, through which all the session signaling information between the UA and the network flows. The PCRF has access to network policies and subscriber related policies. Based on those it can grant or deny bearer (media) requests by the user by setting the appropriate policies in the PCEF.

Border Gateway Function (BGF) – The BGF is a specialized IP router through which all the media flows. It may contain firewall functionality, and also application layer gateway functionality. It can also act as a policy enforcement point by receiving instructions from a policy control function.

Application Servers (AS) – The SIP AS's deliver value-added services (i.e. a service for which a subscriber is willing to pay) and applications in the IMS. They are an optional part of the IMS and may be added independently from each other as operators introduce new services. They interface with the core via SIP and are invoked by the CSCF depending on each subscriber's profile.

Media Resource Control Function (MRFC) and Media Resource Control Processor (MRFP) – The MRFC and MRFP together can provide the traditional telephony functions of detecting and generating tones from and towards the user, and generating recorded announcements, but can also provide advanced media services such as mixing incoming media streams (e.g. for conferencing), and processing media streams (e.g. for transcoding between different codecs). They can be used as part of an overall solution for Voice Mail, or Interactive Voice Resource (IVR) such as those used to direct incoming calls to the right attendant in a call center.

Location Retrieval Function (LRF) – The LRF obtains location information about the mobile subscriber for other services, such as emergency calling and location based applications. The LRF may interface with other functions in the mobile network to obtain this information.

3.3 Reference Points

The IMS standards define strict reference points for signaling between all network elements. There are more than 20 reference points between internal network elements, and between them and external entities (UA's, other networks) [3GP002]. A reference point is an interface. An interface is always a logical interface, i.e. the way two logical functions communicate with each other. Sometimes it is also a physical interface, if the logical units do not reside within the same hardware platform. Strictly speaking, every interface is of interest both to the engineer trying to protect it from attack, and to an attacker. In practice, however, we can limit our work to those interfaces that span two or more logical functions.

A reference point generally specifies not only the application layer protocol, but also the underlying protocols. For example, in the figure 3.2, the layers of the ISC reference point are shown

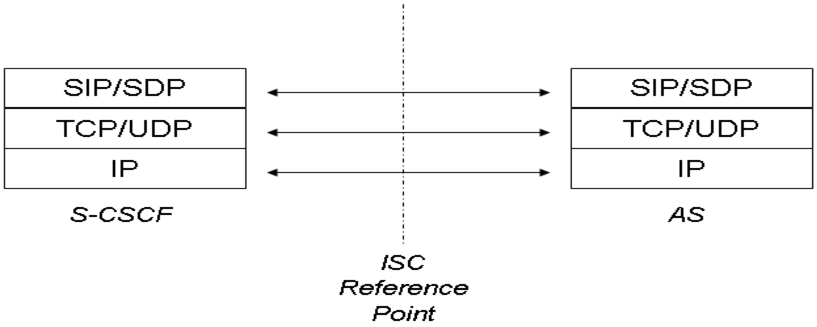


Figure 3-2 - ISC Interface, Protocol Layer Model

SIP/SDP: defined in [RFC3261], [RFC2327], [RFC3266], [RFC3323], [RFC3325], [RFC3262], [RFC3264], [RFC3311], [RFC3312] and [24.229]

TCP/UDP: defined in [RFC793] and [RFC768]

IP: defined in [RFC791] or [RFC2460]

What follows is a brief description of the more critical reference points with respect to the security aspects discussed here:

***ISC* –** Reference point between S-CSCF and Application Servers, uses the SIP protocol as defined in RFC 3261 [ROS02] enhanced with 3GPP specific extensions. This is the interface via which the S-CSCF invokes all the applications the operator has deployed.

Cx – Reference point between S-CSCF and HSS, uses the Diameter protocol as defined in RFC 3588 [CAL03] enhanced with 3GPP specific Attribute Value Pairs (AVP). Via this interface the S-CSCF retrieves and updates information in the IMS database.

Sh – Reference point between HSS and Application Servers, uses the Diameter protocol as defined in RFC 3588 [CAL03] enhanced with 3GPP specific AVP's. This interface is used by applications which need subscriber data in order to perform their functions, or which store *their own* data in the HSS for reliability or redundancy.

Mw – Reference point between CSCF's, uses the SIP protocol as defined in RFC 3261 [ROS02] enhanced with 3GPP specific extensions. This is the main reference point for signaling between the call processing elements within the own IMS and with other IMS networks.

Mr – Reference point between an S-CSCF and an MRFC. uses the SIP protocol as defined
RFC 3261 [ROS02], other relevant RFC's, and additional 3GPP enhancements. The S-CSCF can invoke the playing of announcements or tones towards the subscriber. It is also used for controlling media type features like conference calling, video downloads, etc.

Gm – Reference point between CSCF and UE (User Endpoint). Uses the SIP protocol as defined in RFC3261 [ROS02]. This is the access interface into the IMS core network.

Gx – Reference point between Policy Control Enforcement Point (PCEP) and Policy and Charging Resource Function (PCRF). Uses Diameter per RFC 3588 [CAL03] plus additional 3GPP specific parameters. Used for downloading policies from a policy control function into a border gateway.

Rx – Reference point between PCRF and P-CSCF. Uses Diameter per RFC 3588 [CAL03], plus additional 3GPP parameters. The Proxy-CSCF communicates application layer policy requirements (bandwidth, quality of service) to the policy control function, which validates them and asserts them via Gx towards the border gateway.

Rf – Reference point between the Offline Charging Function and any element providing offline charging information. Uses Diameter per RFC 3588 [CAL03], plus additional 3GPP parameters. The S-CSCF and other elements upload their charging records towards a charging element which does post-processing on those records.

Ro – Reference point between the Online Charging Function and any element providing online charging information. Uses Diameter per RFC 3588 [CAL03], plus additional 3GPP parameters. Online charging, as opposed to offline, is used for pre-paid applications, i.e. the application is disconnected when all credits have been exhausted (e.g. pre-paid phone cards).

3.4 Design Principles

IMS is not the only next generation network architecture, nor the first. It should come as no surprise then, that some of the design goals behind the IMS had already been previously proposed.

One of the initiatives that precede the IMS is the Multiservice Switching Forum (MSF) [MSF00]. Started in 1998 by a consortium of service providers and system suppliers, its stated goal was to develop and promote an open-architecture, multiservice switching systems. Multiservice means that its target application was not just voice (telephony) but anything that could be transported over different packet data technologies. Switching meant that this was to replace the then (and now) still predominant telecommunications technology, Time Division Multiplex (TDM). In its Release 1 document, the MSF presented its goals as those of separating switching systems into clearly defined control plane, switching plane, and adaptation plane.

Another early telephony over IP initiative, the International Softswitch Consortium (ISC) [ISC00], now defunct, also proposed in 1999 “open standards and protocols, and new application development for a distributed set of hardware and software platforms which can seamlessly interconnect the traditional telephone network with information and applications currently available only over the Internet.

Both of these organizations and others preceded 3GPP in trying to define the principles of a Next Generation Network. In fact, before we try to list those, we should specify what makes a network “next generation”. We can define the basic characteristics of an NGN as:

- Open architecture (i.e. standardized non-proprietary interfaces among major, independently procurable and deployable functions)
- Exploit commercially available computing platforms (i.e. non-proprietary hardware and operating systems)
- Separation of media and control. This is just more than common channel signaling (CCS). CCS has to do with signaling “out of band”, i.e. via a separate physical link and using a byte-oriented protocol. But nothing says that both the media and the signaling cannot originate from or terminate to a single network element. By separation of media and control it is also meant that there is a network element, optimized for handling media, and a control element, optimized for understanding communications protocols which commands the media element on how to switch (and transcode, prioritize, block, etc.) the arriving media streams.
- Multiservice, in other words, the network should be able to be used for more than just telephony.

The ITU-T (International Telecommunications Union) has also laid out the general characteristic of an NGN in ITU-T Recommendation Y.2011 [ETS001].

With this concise definition of an NGN, we can now state the goals of the IMS and how it goes beyond a plain NGN.

3.4.1 Separation of Applications, Control, and Media

If the next “next generation network” is to be more flexible, less proprietary, more adaptable, more nimble, in a word, more like the Internet, it’s clear that modularity is essential. It needs to be possible to make a change here and there, or to introduce a new application, or to develop a new type of access, without having to touch a lot of different network elements. The first step in this is to realize that there are three essential parts to communicating: the what (i.e. what’s being sent from one user to another, voice, video, or messages), the how (i.e. what are the features of that communication session), and the where (i.e. how do these two users find each other). These three parts are respectively Media, Application, and Control.

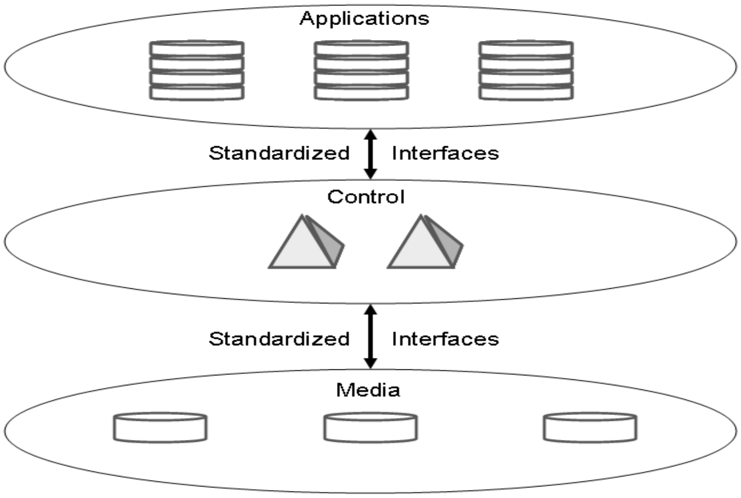


Figure 3-3 - Separation of Functions

The control and the application cannot be interdependent, because the application probably does not care where a user is, what type of access is being used, how much load is on the network, or whether the user also has two other sessions active with different applications.

The application and media cannot be interdependent, because most of the time, it is of no relevance to the application whether the media is IP all the way, or IP part of the route and TDM the other half.

And finally, the control and media also cannot be interdependent because there are many types of media, many routes to a destination and many times when there is one without the other.

In the end, separation of these three parts is about acknowledging that the internet has had an influence on telecommunications that communications are getting more heterogeneous, and recognizing that it must be possible for any future application, if so programmed, to manipulate sessions, content, events, in any way it desires, without constraint, in order to create a richer communications experience. This is what the IMS and an all-IP network aim to provide.

The Multiservice Switching Forum, by not emphasizing the separation of Application from the control layer, has perhaps missed an opportunity to reach the relevance that the IMS is gaining.

3.4.2 Access Network Independence

Another area in which the MSF came up short, in our opinion, is in not recognizing that the future was IP. One does not need to go further than the Release 1 MSF architecture document [34] to see that there was too much written about access to ATM SVC Services, ISDN access to ATM Gateway, Voice over ATM, Frame Relay, etc. The control plane has too many controllers (for IP/MPLS, for SS7, for ATM). Ethernet IP is just one more. In hindsight, admittedly always 20/20, it is clear that an operator which tries to integrate all these technologies and deliver homogeneous services over them, is not going to simplify its architecture but complicate it, and it’s not going to be nimbler and faster to market but be encumbered by them.

Figure 3.4 shows a different structure, the one envisioned by the IMS:

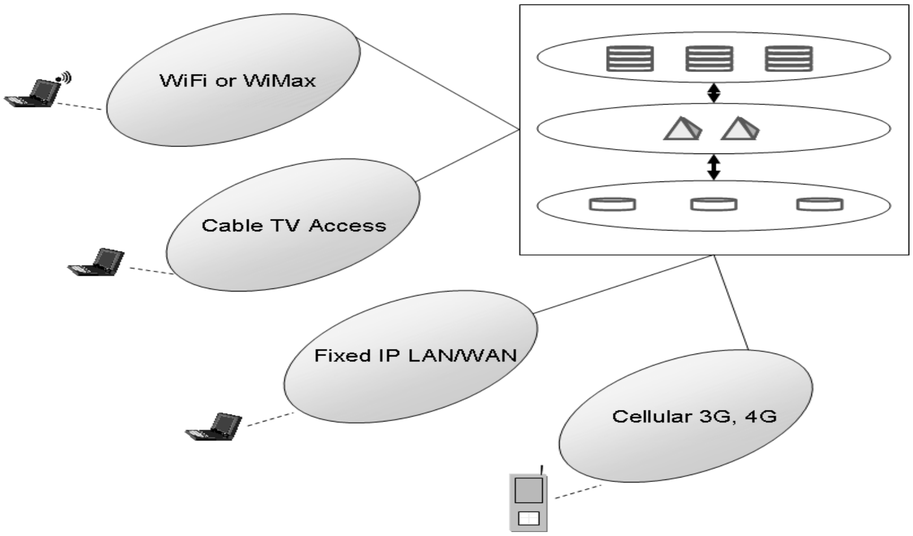


Figure 3-4 - Access Independence

The recognition that eventually all endpoints will be IP and speak SIP, was perhaps the greatest foresight of the IMS developers. Therefore, the goal was to design a control network which left the adaptation to elements at the edge of the network and which could assume that all session requests will come as SIP messages over IP. In this way, the designers can leave it up to vendors to provide adaptation gateways (black phone to SIP, ISDN to SIP, SS7 to SIP, PRI to SIP, H.323 to SIP, and any others), and concentrate on the main goals of the control plane: to build, modify, tear down sessions, and to know when to invoke the right services and applications, while keeping the correct charging records, and guaranteeing the agreed on quality of service.

With contributions from TISPAN, and CableLabs, and others, the IMS today is being deployed in truly access agnostic environments and many times serving subscribers which before would have obtained services from entirely separate core network technologies.

3.4.3 Avoidance of Duplication of Common Resources

A clear drawback of communications networks prior to IMS (although the MSF, in theory would not have shared this problem) was that new applications when able to be deployed over a common network, usually required new and different functions in the areas of Charging records, Security, Subscriber Database, and other areas that are only tangential to the application itself. This is referred to as the "silo" effect, because the application, sitting at the top, requires its own versions of other "enabling" functions like those mentioned above.

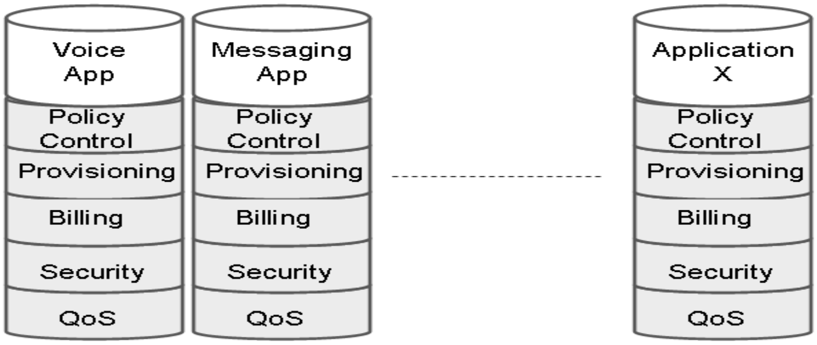


Figure 3-5 - The "Silo" Effect

The IMS aims to do away with the silo effect. By clearly separating applications from the rest of the network via a control interface to the CSCF (the ISC interface, see section 3.3) and via a database access interface to the HSS (the Sh interface), the following benefits arise (see Figure 3.6):

- Common functions like charging, security, policy control, authentication, and quality of service, can be handled, for the most part, in the control plane, for every application. They are deployed once and administered centrally, not as separate functions, with considerable savings in operating expenses.
- Subscriber data can be centralized. Imagine having only one data store for fixed telephony subscribers, cellular service subscribers, business users, cable subscribers, etc. What before would have been separate hardware and software platforms, with different maintenance requirements

(patching, upgrading, expanding, alarming) and different administration interfaces, can now be one single platform

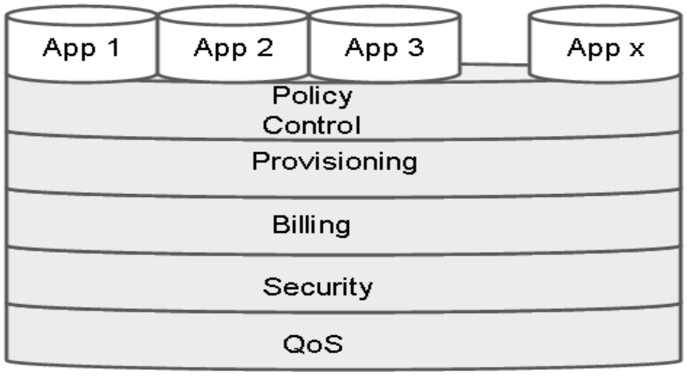


Figure 3-6 - One-time Deployment of Supporting Functions

3.4.4 Re-use of Internet Open Interfaces and Technology

The three previous design goals would not give any great advantage to an operator if it still has to depend on a single supplier for all the components of an NGN. It is no advantage if more applications can be deployed more rapidly by adhering to the above design goals, but the operator does not get the benefit of more competitive prices and a wider selection of components and applications made possible by open interfaces.

Clearly, the idea of open interfaces is not new. In TDM networks there were already open interfaces, or standardized protocols, for central office switches and other elements like Signaling Control Points (SCP) of different manufacturers to be able to interwork. Examples are the Signaling System Number 7 (SS7), and the different parts that were carried by it: Message Transfer Part (MTP), ISDN User Signaling Part (ISUP), Mobility Application Part (MAP), etc. Other protocols like X.25, and obviously the entire family of IP protocols, are also “open”. It is in fact the IMS designers original vision of making the next generation network as successful and open as the Internet, that moved them to re-use as many IETF-defined protocols as possible, including SIP. SIP and the closest competitor at the time it was chosen, the ITU’s

H.323 [ITU06] present a stark example of the difference between an open, user-friendly, understandable (text based) and expandable signaling protocol (SIP), and a closed, complicated, arcane (bit oriented) protocol which only a few telecom engineers would ever master. Like the protocol from which it was derived, HTTP, SIP is guaranteed to be able to be used by millions of internet programmers, and to grow by means of the open process of Requests For Comments (RFC's) which guides the IETF.

SIP is the base of all the session control reference points in IMS. Another protocol gaining in prominence for other control plane functions like charging, policy control, and database access, is Diameter [CAL03], the successor to RADIUS [RIG00].

3.4.5 Decoupling of User-Device Identity

This is a new concept in the area of telecommunications. Currently, a telecom network, be it for fixed communications, cellular, or even for cable TV, does not have a concept of addressing a real person, but rather a device. In other words, if someone dials 561 542 7318, the call comes to a telephone situated at a specific location in Boca Raton, Fl, USA. No matter that the person who owns that destination is there or not (Call Forwarding notwithstanding). If someone dials 770 806 4834, the call comes to a cell phone which may or may not be with its owner in the Atlanta area. And if someone orders a movie from address 1023 Maple Avenue, Cherry Hill, NJ, USA, the bill will come to the owner (renter) of that digital cable box. In none of these cases there exists the possibility of a person other than the original contracting user of the device to "log on" into that device, and have his or her usual services be delivered to it.

An exception to the above has been in the GSM cellular networks, where the Subscriber Identity Module (SIM) resides in a memory chip which can be moved from one cell phone to another, thereby allowing a certain mobility of identity.

This decoupling of user from device has been possible for a long time now with personal computers. With most operating systems for the last ten years, it has been possible, if the network is so set up, to allow any user to log on into any particular PC and instantly “own it” from the point of view of profile of settings and services which are permitted at that workstation. The IMS has set this same goal for telecommunications services by defining the concept of Public Ids, as shown in Figure 3.7. A Public ID (or several), is (are) assigned by an IMS network to each user upon contracting the service. All the user information and data on subscribed applications (service profile) are associated to this Public ID. A user can then “log on” to any IMS device (which has a unique Private User ID), anywhere in the world, and register for service at that location only, or at multiple locations if so desired. In essence, by registering from a particular device, the user is informing the network where he or she can be found in order to receive service.

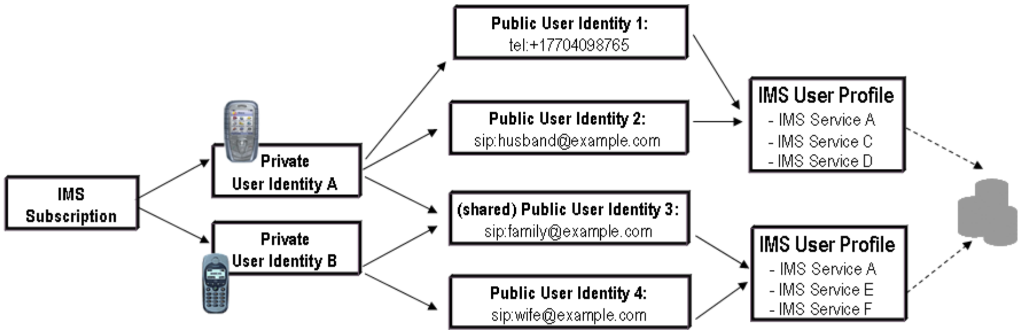

IMS Subscription
Private User Identity A
Private User Identity B
Public User Identity 1: tel:+17704098765
Public User Identity 2: sip:husband@example.com
(shared) Public User Identity 3: sip:family@example.com
Public User Identity 4: sip:wife@example.com
IMS User Profile
- IMS Service A
- IMS Service C
- IMS Service D
IMS User Profile
- IMS Service A
- IMS Service E
- IMS Service F

Figure 3-7 - Private/Public User ID's

This design goal addresses the critical need in telecommunications of the last decade, and years to come: mobility. No longer should a user of services be tied to a street address, or to a particular cell phone, PDA, lap top, or digital cable box. If he/she is on vacation in a time-share five thousand miles from home, it makes sense that some services can be “migrated” to local SIP-capable devices.

3.4.6 Operator Control of Security, QoS, Charging

In the first experiments with Internet telephony and SIP, the idea was to deploy another service which would be free and which would allow a user to bypass the telephone companies, especially for long distance calling. Technically, the goal was to replace the centralized control of telecommunications networks by moving the intelligence to the end devices: the software application residing on a lap top and eventually on a smart phone. The intention of companies like Vonage and Skype is to use the “free” bandwidth available over broadband connections for these services, which are offered on a best-effort basis. In other words, no guaranteed user experience, very little security, and of course, no need for charging.

The 3GPP, being an organization with representatives from large telephone companies, who are in business to make money, and large equipment providers who want to sell network technology, obviously is interested in advancing the state of the art in the technology used by those operators, in order for them to stay in business. It needs to facilitate the introduction of new services for which users will be willing to pay money.

This means that those services will need to have the same or better features as they are used to, in terms of security and predictable quality. The 3GPP has set as one its design goals to ensure that the IMS architecture provides both of them. And it also follows that there must also be a charging component in every part of the architecture where it is needed.

3.5 Some Published IMS Announcements

The IP Multimedia Subsystem architecture has been in development since 2001 and is now reaching its fourth 3GPP release, Release 8 (the first 3GPP release to include IMS was Release 5). Many IMS vendors, mainly Ericsson, Nokia Siemens networks, and Alcatel-Lucent, have meanwhile gone through 4 or more release cycles of their IMS products. As a result, it is the opinion of most experts that the standards and the products that use them are mature.

So why have operators been so slow in deploying networks based on IMS? There is no single answer to this question. Certainly it’s been only within the last 2 years that both the standards and the products have reached this level of maturity. It is also true that it takes a major operator like Verizon or AT&T at least a year of evaluation and testing of a new complex technology like IMS to even be able to make vendor decisions, let alone design and commercially deploy a network. But perhaps a more important reason for the delay has been the lack of need (“killer application”) up to now for the technology, and the fact that the digital TDM systems in place are still providing reliable telephony service.

In the last couple of years, however, we seem to have seen a flurry of decisions for IMS, as listed here, and certainly more are coming. What’s finally driving the adoption of the technology? We believe the top reasons are:

- The discontinuation of manufacture (end of life) and support by major vendors like Alcatel-Lucent, Nokia Siemens, Ericsson, Nortel, of TDM equipment. The consequence is that it is getting more expensive for operators to replace and maintain this infrastructure.

- Acceptance of other forms of VoIP. Early adopters like Vonage, the cable companies, and other Internet providers have helped push VoIP, although not IMS-based, to the point that it is now an accepted and proven technology.

- Widespread availability of broadband, both wired and wireless, over which VoIP and other multimedia services enabled by IMS can be delivered.

- Realization that current VoIP technology is not future proof

As can be seen from the following compilation of announced decisions for IMS, some of the major telecommunications operators are finally making their decisions for IMS, and in some cases, for specific vendors. We believe that with the telecom giants leading the way, the critical mass necessary is being reached for wide IMS acceptance and deployment. The list below is not a complete list of all IMS decisions world-wide.

3.5.1 Verizon

Just in February of 2009, after more than a year of network trials which followed an open tender process among the major IMS vendors, Verizon has finally announced the selection of Nokia Siemens Networks and Alcatel-Lucent as their two IMS vendors, at the GSM World Congress in Barcelona [VER09]. This announcement came at the same time as Verizon's decision on their LTE 4G deployment choice for North America. It is significant that both decisions have been made in concert: as this work hopes to prove, only an architecture like IMS can control the multitude of applications that Verizon will be able to provide to wireless users over 4G broadband (where Verizon demonstrated download rates of 50 to 60 Mbps peak speeds in the 700 MHz spectrum). In addition, Verizon will also be able to serve fixed line and fiber connection subscribers with the same IMS control core. Verizon expects to start offering commercial LTE-based service in the United States starting in 2010.

Verizon gave the following rationale for selecting an IMS control core together with an LTE access network:

- A vision to provide ubiquitous global wireless broadband connectivity and mobility.
- Consumer demand for mobilizing the many applications they frequently use when tethered to high bandwidth wired networks.
- Enabling rich multimedia applications regardless of access technology; goal to offer converged applications and services on its wireless *and* landline broadband networks.

Verizon also disclosed that the company's overall spending program ($17 Billion in 2008) will be shifting from older technologies to new strategic initiatives, such as LTE and IMS and that they will be creating the Verizon LTE Innovation Center in Boston, with the mission of being the catalyst for development of non-traditional products for use on LTE networks.

3.5.2 KPN

Royal KPN NV, based in The Netherlands, announced in 2007 that they would start IP Multimedia Subsystem-based (IMS) voice services to their broadband subscribers. KPN planned to spend from 1 to 1.5 billion Euros on its all-IP network, to be completed by 2010 [KPN07].

KPN is the leading provider of telecommunications services in the Netherlands, serving customers with wireline and wireless telephony, Internet and TV services. To business customers, KPN delivers voice, Internet and data services as well as fully-managed, outsourced ICT solutions [KPN08].

The reason for the decision in favor of IMS was to "bring new IP high bandwidth broadband services to the customer and switch off legacy networks." KPN's network consists of VDSL and Fiber To The Home (FTTH). After voice services, KPN plans to offer other types of communication services such as IP Centrex, and other mobile as well as fixed services.

Among other reasons for the decision in favor of IMS, KPN mentions:

- Savings of hundreds of millions of Euros a year in reduced network maintenance costs, which includes the replacement of existing SS7 technology [KPN08].
- The ability to introduce new services rapidly.

3.5.3 Telia Sonera

Telia Sonera, the Finnish telecom operator, announced in May 2007 that they had made their decision to deploy IMS to offer IP-based services such as VoIP, video calling and instant messaging. Telia Sonera chose Nokia Siemens Networks following the usual long and extensive technical evaluations [TEL07].

Telia Sonera waited until it was convinced that IMS would interoperate successfully with other SIP networks, and would function smoothly across national boundaries from day one.

Additional features that Telia Sonera will eventually deploy over IMS,aside from voice calls, messaging, and viewing videos and photos, is network-based phone book which allows users to have the same address book on their fixed phone, mobile and broadband terminals, and can see whether the person they want to contact is offline or busy.

3.5.4 AT&T

AT&T, and previously, Cingular Wireless (now a part of AT&T), has made several forays into VoIP and IMS but has not really yet jumped in with both feet. However, it’s clear from their published document describing the CARTS architecture described below (CARTS stands for Common Architecture for Real Time Services), that their sights are set on IMS as the control core.

One of the services AT&T has made available with IMS as its control core, albeit in limited markets, is U-verseSM Voice. It is a voice over IP service for consumers which should eventually integrate wireline and wireless voice, with broadband and TV services [ATT07]

U-verse Voice also includes standard calling features like caller ID, click-to-call, a unified mailbox for wired and wireless messages, and an online management portal. AT&T also allows wireless customers who subscribe to an AT&T Unity Worldwide Calling plan to call any AT&T U-verse Voice number without using up their wireless minutes [ATT07]. Alcatel-Lucent provides the IMS components.

But AT&T also continues to market is non-IMS-based CallVantage VoIP service to customers who don’t buy U-verse video services, particularly customers who live outside AT&T’s local phone service territory.

Eventually, AT&T plans to deploy CARTS to unify its existing incompatible next generation services under an access agnostic and multi-aplication capable IMS control core.

CARTS is advertised as [ATT08]:

“IP Driving Anytime, Anywhere Communications”; the network able to deliver applications and information anytime, anywhere, and to any IP-enabled device.

A full range of information and content which will be accessible via a single device, and which will be delivered by the network over the best available connection at a given place and time.

The architecture that will enable AT&T to build intelligence into its network and share information with any of the “three screens” – the PC, TV and wireless device.

AT&T had plans to begin introducing CARTS-enabled applications for residential and business customers in 2007 and 2008. Obviously this has not happened yet. It is thought that AT&T will start an evaluation period for CARTS suppliers in 2009. Once CARTS is launched, some examples of the applications planned are: Video Share, VoIP service, VoIP services and applications for enterprise customers, long distance phone network migration to IP, social networking, music, location-based service enabler, TV voicemail, TV talking caller ID, TV wireless caller ID, and dual-mode phone.

3.5.5 China Telecom

China Telecom, an operator with more than 460 million subscribers, is using an IMS core network to provide its video monitoring solution [CHI08]. This solution uses a multi-media service delivery platform to deliver to subscribers high-end video for remote monitoring over broadband, which in subsequent versions will include mobile devices.

3.5.6 Chungwha Telecom

Chungwha Telecom, Taiwan’s provider of fixed, mobile, and internet services, with a subscriber base of about 8.8 million, has embarked on a five-year plan to deploy a next generation network centered around IMS, for voice over IP services [CHU07]. This will include the migration of the current “legacy”, i.e. non-IMS, VoIP architecture to a 3GPP-compliant framework.

3.5.7 Com Hem

Com Hem, the leading supplier of triple play services over cable in Sweden (TV, Broadband, and Telephony), with around 40 percent of all Swedish homes connected, has introduced an IMS platform for commercial delivery of voice over IP services, and other applications [COM07]. Com Hem mentioned in its decision to deploy IMS the reduction of operations expenses and the possibility to introduce new telephony services and applications quickly.

3.5.8 Vodafone

UK-based Vodafone Group Plc, which bills itself as the world’s leading mobile telecommunications company, and which has a 40% stake in North America’s Verizon Wireless, announced in July of 2007 that it had signed a contract with Ericsson for the purchase of IMS equipment for its subsidiaries in Germany and Portugal [VOD07]. Vodafone adopted IMS to enable its strategy of combining in its service offering “the best of the mobile and internet/PC worlds”.

3.5.9 Telefonica

Telefonica, the Spain-based provider of telecommunications services, present in 25 countries and and with about 252 million customer accesses, 188 million of those mobile subscribers, and with high numbers as well of Internet and pay TV users, announced in February of 2007 that it had signed Alcatel-Lucent for providing it with an IMS architecture for its Presence Server project [TE707]. This is a service by which its subscribers can manage their incoming communications, based on originator, time of day, location, or any other information relevant to the call. In addition, commuincation services such as instant messaging, video mailbox, push-to-talk, and others, can be used by subscribers to the service in both wired and wireless environments.

3.5.10 North American Cable Companies (MSO's)

The author knows that three large MSO's in North America have made IMS vendor selections and are currently in the process of designing or building their next generation multimedia services networks around this IMS core. Since these companies have not yet made this information public, the details cannot be presented here.

Now that we have introduced IMS, we will do a short survey in Chapter 4 of other NGN frameworks which have been developed, before trying to develop the concept of an abstract NGN in Chapter 5.

4 OTHER NEXT GENERATION NETWORKS

As previously mentioned, there are other Next Generation Networks which have in the last years adopted many of the design goals as the IMS. We will mention here three in particular:

- CableLabs PacketCable 2.0
- TISPAN 1.0
- MSF 3.0

4.1 CableLabs PacketCable 2.0

We describe here the architecture being defined by the consortium of North-American Multi Service Operators (MSO), better known as the cable television companies. CableLabs is the standardization and conformance body founded in 1988 by a group of North American cable companies, to do research and development on new cable telecommunications technologies. PacketCable, is the CableLabs initiative designed to develop the necessary specifications for enabling advanced voice and multimedia applications over the high speed cable plant. PacketCable 2.0, unlike its predecessor, PacketCable 1.5, has adopted 3GPP IMS as a basis for its control core network and adds those building blocks necessary for implementing services in a digital cable environment. Consequently, CableLabs has also therefore selected SIP as the session signaling control protocol in the core and for the user end-points, which will need to be upgraded from their current NCS (Network Control Signaling) signaling. PacketCable 2.0 defines the set of specifications that govern the architecture, introduce the logical functions, and set the protocols for their interworking. They can be found at [CAB00] and the architecture is defined in [PAC08].

4.1.1 PacketCable 2.0 Architecture

A diagram of the main building blocks can be seen in Figure 4.1.

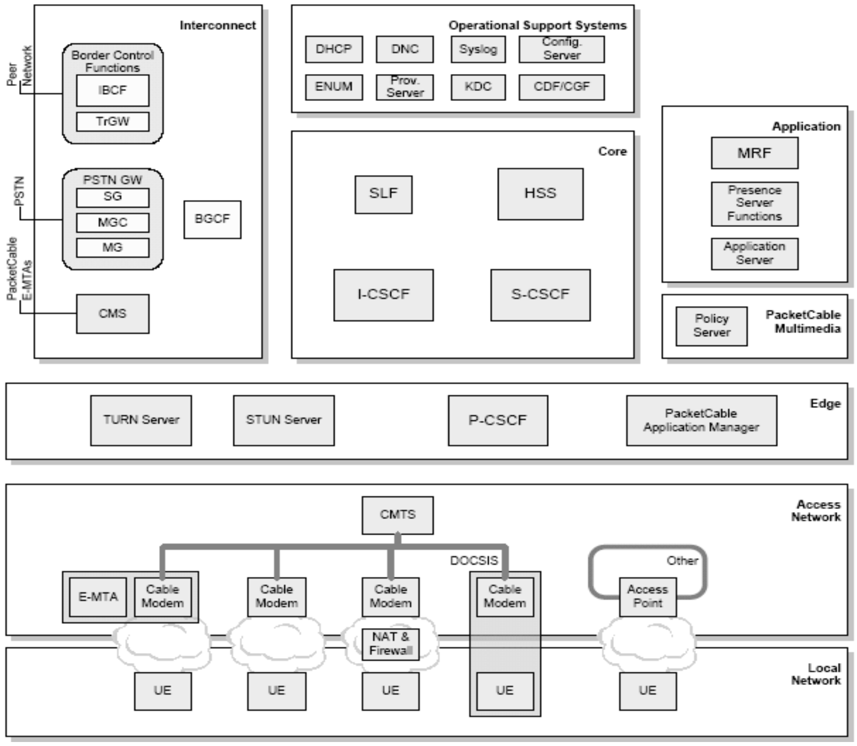


Figure 4-1 - PacketCable 2.0 Architecture

Figure 4.1 shows the core of the network is comprised of the two main functions in the IMS architecture: the *Home Subscriber Server (HSS)* and the *Call/Session Control Function* in its three variants (I/S/P). A *Subscription Location Function (SLF)*, not described in the previous function, but also a network element defined in the 3GPP IMS, is shown in the core domain and can also be deployed although this will be rare in most networks. The SLF can be queried by the I-CSCF upon first registration by a subscriber, to determine in which HSS its subscription information can be found, when there is more than one HSS deployed. Given the almost unlimited capacity of most vendors HSS's, it is not likely that SLF's will need to be deployed. Another case in which an SLF may be needed is if an operator has two different HSS's from two different vendors. The Proxy CSCF is also part of this architecture, having the same functionality as in the IMS and placed here in the edge domain.

Other key functions in a PacketCable 2.0 network are:

The TURN and STUN servers – These network elements are used for traversing Network Address Translators (NAT). A NAT is usually used within the customer premises of a user in order to be able to use one single IP address for several devices. The NAT translates between the internal, non-globally routable IP address in the home and the external globally routable address. When this is done, some protocols such as RTP, which embed IP addresses within the payload will no longer function. TURN and STUN servers solve this problem by discovering what the external facing address is, and using that instead in such protocols. They consist of a client-server configuration, the client being within the end-device needing to discover its external address beyond the NAT, and the server being within the operator's domain. In brief, the STUN/TURN client will send a message to the STUN/TURN server which will see it after its origin address has been "NATted" i.e. translated into the global address. It will then insert that address into a response message and send it back. STUN stands for Simple Traversal of UDP over NAT's, and has been specified in RFC's 3489, and later 5389 [ROS08]. TURN stands for Traversal Using Relay NAT, and is still in Internet draft stage.

The PacketCable Application Manager (PCAM) and Policy Server (PS) – The PCAM and PS were introduced in the Packet Cable Multimedia architecture [PACM] which is a predecessor to PacketCable 2.0. They are two separate logical functions which together are responsible for translating application level quality of service resource requirements into specific media policies to be installed on the CMTS for enforcement.

The PCAM is an entity which resides in the application domain, i.e. the logical part of the network which contains the elements that offer applications and content to service subscribers. The PCAM defines service policies and couples subscriber-initiated requests for content and services with the network resources needed to meet those requests. It is also its function to authenticate and authorize the client requests. Once this is done, it translates the client requests into specific network resources and sends a request for these resources to the Policy Server.

The PS performs the function of the policy decision point (PDP). It receives policy requests from the AM and applies those policy rules that have been defined by the operator before forwarding the request to the CMTS (see below). The PS takes the application requested QoS requirements and translates them into gate commands to the CMTS after modifying them to account for dynamic parameters like available resources, security considerations, time-of-day, etc.

Cable Modem Termination Server (CMTS) – As the name implies, the CMTS terminates the Cable Modem (CM) connections at the operator side. Cable Modems are used for modulation/demodulation of the signals that transport data to and from the customer premises. Both the cable modem (CM) for upstream traffic, and the Cable Modem Termination System (CMTS) for downstream traffic provide also Quality of Service (QoS) by shaping, policing, and prioritizing traffic according to the QoS Parameters defined by the operator.

Presence Server Functions – Like in the 3GPP architecture, the PacketCable Presence Server has the function of keeping, updating, and sharing information about a subscriber’s availability (online/offline), willingness, mood, and other qualifiers as appropriate. This information is managed by the users and it is the user who can give or withhold access rights to it by other users.

Operational Support Systems – Like any other IP-based network, a PacketCable architecture requires essential functions for IP address allocation, translation of Fully Qualified Domain Names (FQDN) into IP address, key distribution for symmetric cryptography, etc. Some of these functions are provided by Dynamic Host Configuration Protocol (DHCP), Domain Name (DNS), and Key Distribution Center (KDC) servers respectively.

Application Servers – These servers may perform a multitude of functions in a PacketCable network. Perhaps the most important is the Voice Application Server. PacketCable has defined the requirements for residential telephony service for SIP end points in a separate specification [PACRST]. Other services specified under the same project (PacketCable Applications) are: PacketCable Cellular Integration and PacketCable Business SIP Services.

Other building blocks in Figure 4.1 which are also part of the IMS architecture: the *Border Control Functions, PSTN GW, Media Resource Function, and Border Gateway Control Function.* In the Access and Local networks, the Cable Modem is the user premises element to which IP devices are connected, i.e. it provides the DOCSIS to Ethernet connectivity, and the User Endpoint (UE) is the SIP terminal.

4.2 TISPAN

The Telecommunication and Internet converged Services and Protocols for Advanced Networking (TISPAN) is a standards organization within the European Telecommunications Standards Institute (ETSI) formed to converge the evolution of wireless, wireline, and internet communications networks. TISPAN builds on the work already done by 3GPP on the IP Multimedia Subsystem (IMS), an architecture based on the IETF’s Session Initiation Protocol (SIP).

The goal of TISPAN, is to define the architecture of the Next Generation Network (NGN), via the definition of requirements, frameworks, protocols, and the publication of such via Technical Specifications and Standards Documents.

The NGN as defined by TISPAN will be [ETS001]:

- A multi-service multi-protocol, multi-access, IP based network - secure, reliable and trusted
    - Multi-services: delivered by a common QoS enabled core network

  - Multi-access: diverse access networks; fixed and mobile terminals
  - Not one network, but different networks that interoperate seamlessly
- An enabler for Service Providers to offer
  - real-time and non real-time, communication services
  - between peers, or in a client-server configuration
- Nomadicity and Mobility
  - of both users and devices
  - intra- and inter-Network Domains, eventually between Fixed and Mobile networks
- "My communications services" always reachable, everywhere, using any terminal

NGN Release 1 was launched by TISPAN in December 2005. TISPAN released the Release 2 NGN Functional Architecture in March 2008, with a focus on enhanced mobility, new services and content delivery with improved security and network management. The Release 3 Functional Architecture is planned to be ready for approval in August of 2009.

The primary emphasis for Release 1 were the functions to support real time conversational services such as voice and video-telephony, messaging, presence management, video on demand, video streaming, and the migration of plain telephone services towards an NGN. The primary access for these services in Release 1 is DSL (Digital Subscriber Line), as the overarching goal is the convergence of the core networks serving DSL subscribers (fixed operators) and cellular subscribers (mobile operators).

The current TISPAN architecture, Release 2, is shown in Figure 4.2.

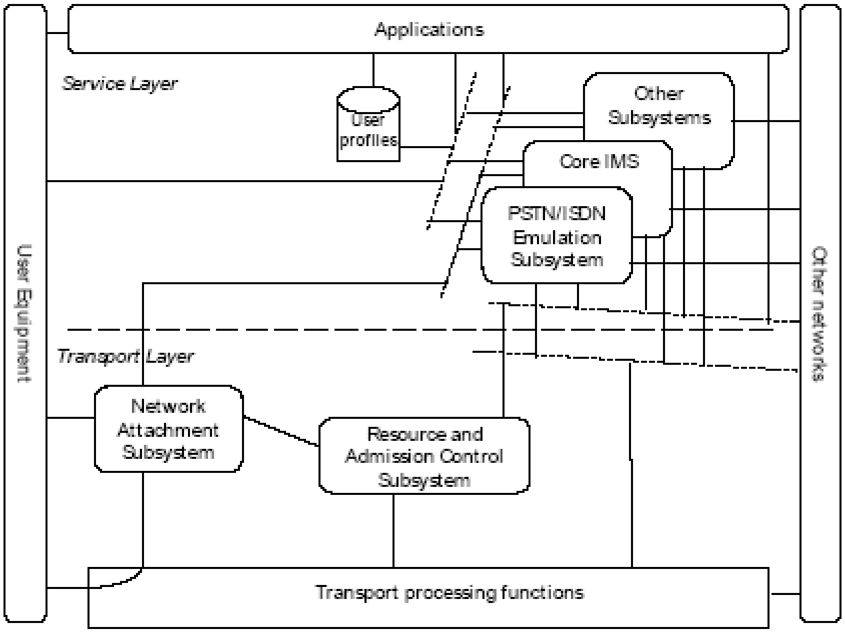

Applications
Service Layer
User profiles
Other Subsystems
Core IMS
PSTN/ISDN Emulation Subsystem
User Equipment
Other networks
Transport Layer
Network Attachment Subsystem
Resource and Admission Control Subsystem
Transport processing functions

Figure 4-2 - TISPAN Release 2 Architecture

The TISPAN architecture is more overarching than the IMS, in that it tries to remain open for the possibility of adding new “subsystems” as demands for new services grows. What TISPAN means by subsystem will be made clear below. Since the IMS itself has been made part of the overall TISPAN architecture, it is to be expected that TISPAN will be even more complicated than the former. It is also probable that no operator will ever implement a real network which uses all of the functions and subsystems defined.

The architecture is built around two layers: a service layer and an IP-based transport layer. These layers must not be confused with the traditional OSI layers, where there was a clear hierarchy of ascending complexity. In TISPAN, the transport layer simply comprises the media control functions, some of them already defined in the IMS (*Media Gateway Function, Border Gateway Function, Media Resource Function*) as well as some new ones.

The service layer is further divided into subsystems, depending on the type of services to be provided via the transport layer, which hides the access technology from the service layer. For example, up to Release 2, the following subsystems have been defined:

The “core” **IP Multimedia subsystem**: provides services (ultimately applications, like voice, video, messaging, etc.) to NGN terminals (i.e. intelligent IP-based end devices) which contain a SIP client.

The **PSTN Emulation subsystem**: provides plain old telephone service (i.e. it emulates a traditional Class 5 digital exchange and the services it provides) to legacy terminals (black phone, fax machine, PBX) which are connected to the IP network by means of TDM-IP gateways.

The **IPTV subsystem**: provides video on demand and video broadcast services.

Each one of these subsystems within the service layer, is defined within its own TISPAN recommendation specification. A more detailed description is beyond the scope of this study. The interested reader is referred to the respective specifications [ETS007, ETS002, ETS012, ETS028] for more information.

4.3 Multiservice Switching Forum

The idea behind the MSF was to create an architecture which permitted the re-use of a common set of switching resources by an “ever changing set of services”. It attempted to achieve this via the separation of control functions from the switching elements (i.e. where the media or data flows through), and from the adaptation elements (i.e. those network elements which adapt access to the switching layer among different types of access like TDM, DSL, ATM, IP, etc). In essence then, the MSF created a control plane, a switching plane, and an adaptation plane. The interfaces between these separate planes where to be precisely defined via Implementation Agreements (IA). It would then be possible for an operator to select from “best of breed” providers for each of the different functions, something impractical in most scenarios up to that point, in order to promote competition. Competition in technology should bring about lower costs and faster innovation.

The MSF left open whether the applications themselves could reside within the control plane our above it. In Release 1 the interface between those two was not addressed. It was clear from the Release 1 specification however, that multiple protocols or API’s could be defined between these two layers [MSF03], or existing ones used where already available, MSF Release 2 specified additional IA’s among all defined functions and elements. It also went from defining just a logical architecture to defining a “reference” architecture, using components for the most part available as commercial products [MSF05]. In MSF Release 3 we see an acknowledgement of the ” reality of wireless-wireline interworking by taking account of the 3GPP IP Multimedia System (IMS) architecture in the core network.” [MSF06]. The IMS architecture is integrated into the core, relegating previous functions/elements from releases 1 and 2 into the background. Also the reality of IP as the de-facto network layer technology is acknowledged by dropping any mention of ATM, and Frame Relay.

A diagram of the MSF Release 3 architecture is shown in Figure 4.3.

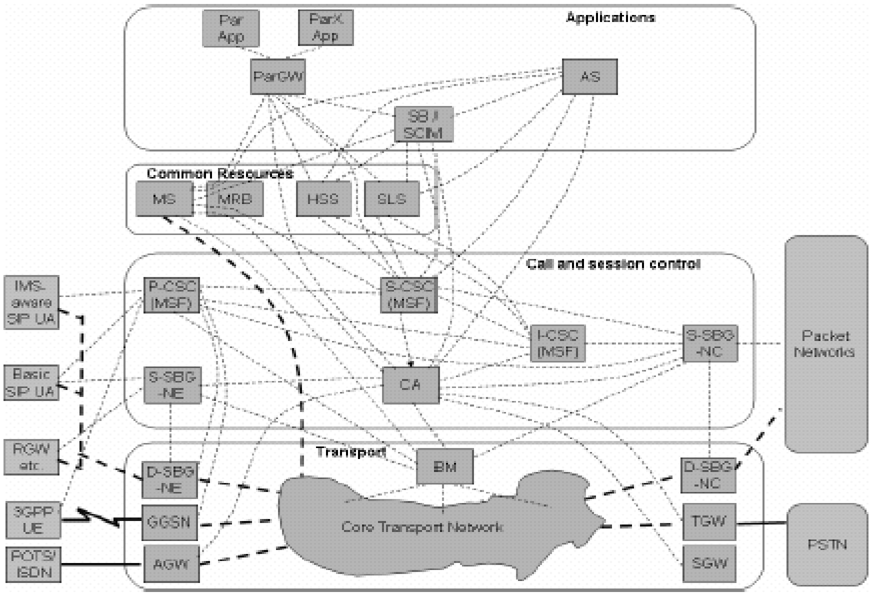


Figure 4-3 - MSF Release 3.0 Architecture

In essence there is very little that separates now the MSF architecture from the IMS. The 3 layers of the IMS architecture have been adopted: *Applications, Call and Session Control, and Transport.* Even the main control functions in the 3GPP IMS have been adopted by the MSF. There is an identical separation of layers or blocks, with the exception of the Common Resources block in which the MSF has decided to gather the database functions and the media resource servers. Most of the differences are only cosmetic as in the renaming of some of the functions. The tasks they perform are the same. It is also the case that the protocols used to interconnect them are for the most part, derived from the equivalent IMS reference point, i.e. SIP and Diameter. It is therefore not necessary to list the functions in the diagram nor to list the tasks performed by each. The reader is referred to the specification [MSF06].

In closing, when we look at other initiatives to define what a next generation network should look like, we see three architectures which have coalesced around the IP Multimedia Subsystem, as is the case with PacketCable 2.0 and the MSF, or which include the IMS as one of its most important components, as in the case of TISPAN.

We now want to take a look at the components of an IMS network, and see what generalizations can be made about them with respect to their placement within the core network, and their security vulnerabilities and requirements.

5 ABSTRACTING THE NETWORK

If one examines the four architectures introduced in the previous chapters: the 3GPP IMS, CableLabs PacketCable 2.0, ETSI TISPAN Release 2, and MSF 3.0,it can be argued that the network elements or functions which comprise them can be classified into one of separate types, based on the purpose they serve *at an application level.* By application level we mean the following. It is clear that all network elements are computers, which communicate with each other and make some kind of decisions based on the data they receive or is provisioned into them. At this level, there are very few distinctions, if any. However, at the application level, the level above physical, data link, network, and transport, to use the OSI layers, these elements do carry out very different functions, and therefore, in the end, may have separate security requirements. We try to identify the characteristics of this abstract network here.

If one can abstract four different architectures into one which represents all of them, the question becomes: can one then apply a security architecture to the generic NGN (the abstract) and have a reasonable expectation that the tools, algorithms, and practices used, would also apply to the real life network they represent?. We will try to answer this question in Chapter 10.

5.1 Functions in IP-based Communications Networks

With the foregoing in mind, we will classify NGN network elements in the following way for the purpose of abstracting the network:

Control Servers – Their primary function is a) the implementation of some specific decision logic which provides a service to the end-user or to the network itself, and b) the routing of control messages which enforce said logic. Examples are: CSCF, Voice Application Server, Push-to-Talk Server, etc.

Database Servers – Their primary function is to hold and disseminate data (upon request or autonomously) on which Control Servers exercise their logic. The data can be subscriber, intra-network, or extra-network related. Examples are: HSS, AAA Server, ENUM Server, SLF, etc.

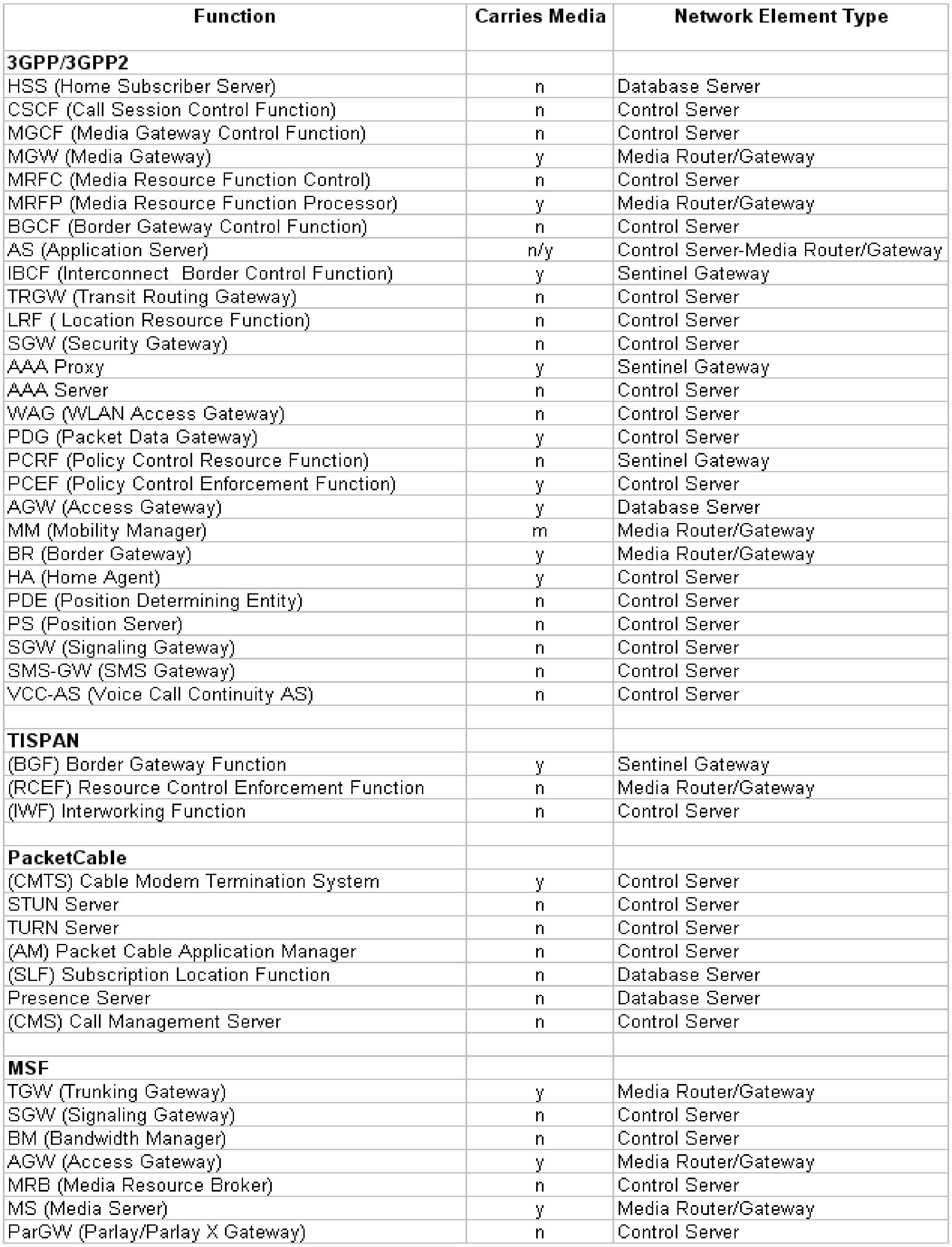
Media Router/Gateway – Their primary function is the “switching” of media in the form of RTP packets. The media can be encoded in different ways (G.711, G.729, etc.) and it can be transcoded (i.e. changed from one type of encoding to another one) as it passes through, therefore the term “gateway”. It may also be transcoded into a non-IP format, for example into PCM in a TDM T1 trunk.

And lastly, what we in this work will call *Sentinel Gateway*s – Their primary function is to protect a network or a network zone by enforcing specific rules for traffic entering the zone, at one or several levels. Network elements falling into this category are usually dedicated special purpose computers, designed for fast, wire speed operation, and also containing deep packet inspection capability.

5.2 Network Element Classification

All elements in the four architectures described in section 2 can be classified as having one of the primary characteristics defined above, i.e. their main function (if not sole function) is either as a Control server, providing a specific value added service or doing control message routing; a Database server, holding information to be used by Control servers; a Media Router/Gateway, transporting media, either transparently or transcoded; or as a Sentinel Gateway, examining control messages and or media crossing the trusted boundaries and applying security or SLA policies.

In table 5.1 we list the most important functions in the four architectures under study, and classify them as one of the network element types defined above. We have included under the 3GPP entries, some which have been defined by the 3GPP variant for CDMA networks 3GPP2.



Function	Carries Media	Network Element Type
3GPP/3GPP2		
HSS (Home Subscriber Server)	n	Database Server
CSCF (Call Session Control Function)	n	Control Server
MGCF (Media Gateway Control Function)	n	Control Server
MGW (Media Gateway)	y	Media Router/Gateway
MRFC (Media Resource Function Control)	n	Control Server
MRFP (Media Resource Function Processor)	y	Media Router/Gateway
BGCF (Border Gateway Control Function)	n	Control Server
AS (Application Server)	n/y	Control Server-Media Router/Gateway
IBCF (Interconnect Border Control Function)	y	Sentinel Gateway
TRGW (Transit Routing Gateway)	n	Control Server
LRF (Location Resource Function)	n	Control Server
SGW (Security Gateway)	n	Control Server
AAA Proxy	y	Sentinel Gateway
AAA Server	n	Control Server
WAG (WLAN Access Gateway)	n	Control Server
PDG (Packet Data Gateway)	y	Control Server
PCRF (Policy Control Resource Function)	n	Sentinel Gateway
PCEF (Policy Control Enforcement Function)	y	Control Server
AGW (Access Gateway)	y	Database Server
MM (Mobility Manager)	m	Media Router/Gateway
BR (Border Gateway)	y	Media Router/Gateway
HA (Home Agent)	y	Control Server
PDE (Position Determining Entity)	n	Control Server
PS (Position Server)	n	Control Server
SGW (Signaling Gateway)	n	Control Server
SMS-GW (SMS Gateway)	n	Control Server
VCC-AS (Voice Call Continuity AS)	n	Control Server
TISPAN		
(BGF) Border Gateway Function	y	Sentinel Gateway
(RCEF) Resource Control Enforcement Function	n	Media Router/Gateway
(IWF) Interworking Function	n	Control Server
PacketCable		
(CMTS) Cable Modem Termination System	y	Control Server
STUN Server	n	Control Server
TURN Server	n	Control Server
(AM) Packet Cable Application Manager	n	Control Server
(SLF) Subscription Location Function	n	Database Server
Presence Server	n	Database Server
(CMS) Call Management Server	n	Control Server
MSF		
TGW (Trunking Gateway)	y	Media Router/Gateway
SGW (Signaling Gateway)	n	Control Server
BM (Bandwidth Manager)	n	Control Server
AGW (Access Gateway)	y	Media Router/Gateway
MRB (Media Resource Broker)	n	Control Server
MS (Media Server)	y	Media Router/Gateway
ParGW (Parlay/Parlay X Gateway)	n	Control Server

Table 5-1: Network Element Classification

The first column contains a list of network elements defined in one of the four NGN architectures presented. Although the list is not exhaustive, it can be seen that some of the elements appear under more than one of architectures, sometimes under a slightly different name. This is partial proof of the impact that the 3GPP architecture has had on other, some times pre-existing, architectures.

The second column specifies whether the network element is used to handle media (i.e. voice or video) or not. There are some elements, like Application Servers, which may or may not handle media, depending on the service provided. As we can see, there is a majority of NE's which do not handle media. They are mainly for signaling, or for controlling the media path or other elements.

Finally, the third column assigns to each network element one of the classes defined above: *control server, database server*, *media router/gateway, or sentinel*.

5.3 NE Placement and Vulnerabilities

Having defined four categories into which every possible network element can be placed, it is of interest to examine the relationships among them, and their placement within the network, in order to try to assess their vulnerabilities, and therefore the correct security patterns to use to defend against them. Security Patterns and Patterns in general, are introduced in Chapter 8.

Figure 5.1 below shows the typical placement in a telecom operator's domain of the each of the four classes of network elements. Dashed lines signify control relationships, solid lines signify media paths, bidirectional arrows show the data flow, and the dotted line signifies the domain that must be made secure from external attacks. Control relationships may signify hierarchical control, i.e. one control element commands the actions of another one, or it may signify simply transport of control information. Each building block may represent multiple instances of the function, and there may be instances where building blocks of the same type communicate or control each other, represented by the looped line on the control/application server block.

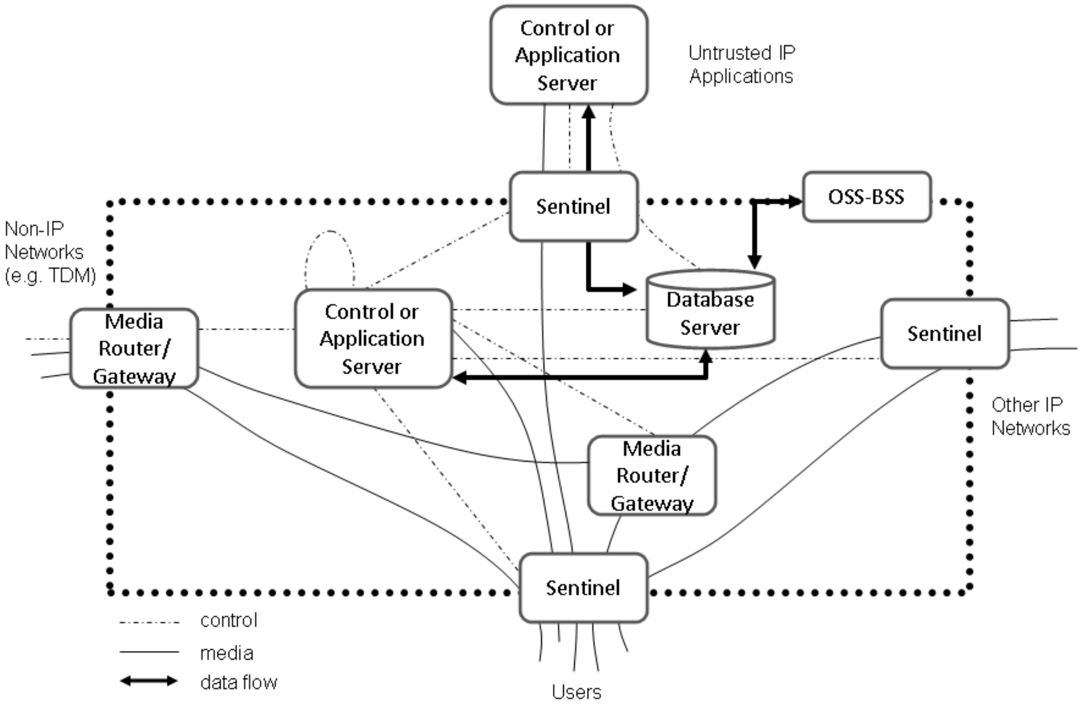


Figure 5-1 - Abstracted NGN Architecture

Some observations can be readily made on this generalized architecture about each of the network element types and the likely threats:

Database servers - There is one network element which has an unambiguous position in the network, the database server. Clearly, this element will always be placed in the most secure place in the network and will be accessed only by a selected group of control/application servers. When, as we will see below, the control/application server (most likely application) is located outside the operator's trusted domain, access to the database server will only be possible via a sentinel network element. Notice that we have included in the picture a class of network elements not considered in our analysis up to now. They are commonly referred to as Operator Services Systems/Business Services Systems (OSS-BSS), and are elements which are not directly used for providing subscriber services or applications, but are needed for other "non-real-time" functions such as: billing, provisioning, analysis, element management, network management, etc. These elements, which may be inside the "secure domain" or not, will need to be considered in the security design in Chapter 10, but are not part of an IMS *core* network and therefore not considered here. They are shown in the diagram in order to complete the shown data flow.

Examples: the most typical representative elements in this category in a next generation network would be the HSS, which contains the subscriber data in IMS; the AAA server, which contains subscriber data in WLAN and other access networks. Another good example is a 3GPP or PacketCable Presence Server, which contains mostly dynamic information about subscriber status.

Likely threats: Most of the time, the database server will be the target of two specific types of attack: information theft or corruption, or more formally, confidentiality attacks and data integrity attacks. More indirectly, database servers can also be targets of Denial of Service. We say indirectly because database servers should not really be visible from outside the trust domain, but the applications that use them could be.

Control/application servers - Next, control/application servers will mostly reside within secure areas of the network. An exception can be when an operator offers services to its users which the operator itself does not control. An example may be as when a cable operator, for example, grants a third party application provider, let's say a specialized service provider, access to its end-user population. The third party provider provides a service to the cable operator's users on a subscription basis, and pays the cable operator a per-user fee. In this scenario, the third party provider must be able to access the subscriber database, and the control elements. But the operator must allow this only via a sentinel.

Examples: Control and Application servers are where most of the service logic resides. The CSCF provides the SIP routing capabilities in IMS, the VAS the voice application, gaming servers the games, PCRF the policy control, and the SGW the signaling translation between SIP and SS7.

Likely threats: control/application servers are the target of usually only one type of attacks: denial of service, or sabotage, where the former is a specialized example of the latter. They interface for the most part with each other only, although some control servers need to control the media router/gateways, and need to communicate with the end users themselves, and with peer networks. Ideally, this interaction with peer networks and users will only take place via sentinel gateways.

Media routers/gateways - The media router/gateway is often at the edges of the network, where media needs to be transcoded or flow controlled or simply routed from one wire to the next. By design, the most common type of media routers are for the latter purpose. Media gateways most often serve the purpose of bridging the two worlds of IP voice communications and the legacy TDM networks.

Examples: IP routers, the MGW function in IMS via which TDM is transcoded into IP and vice-versa, the access and interconnect border functions (A-BGF, I-BGF) also in IMS used for allowing and blocking media streams into the operators domain, and the MRCP, used for injecting tones, music, and announcements into voice calls and multimedia sessions. These network elements interact (are controlled by) control servers for directing/shaping/policing media flow and or trans-coding instructions. Media routers/gateways are also the ideal place for executing functions like deep packet inspection, for example if an operator wants to know what types of applications are consuming the most bandwidth, or wants to monitor if illegal content download is taking place. They are also the logical place where to divert traffic to law enforcement agencies for LI/CALEA compliance.

Likely attacks: media gateways are often employed at the interface border between a TDM network and an NGN. This fact almost guarantees that no attacks will come from one of its sides: the TDM side. This of course applies only to attacks over the interfaces. This network element, like all others can at any time be the target of an insider attack, someone who has physical or management access to the unit. For gateways without a TDM side, if attacked, they will not usually be the direct object of attacks but can obviously suffer the consequences, especially of a DoS attack directed at a network element beyond. Media router/gateway elements interface mostly with other network elements of the same type in peer networks, or with end users themselves.

Sentinel Gateways - Sentinel Gateways are always at the edge of the network, since their function is to keep dangerous or *illegal* control messages and data out. They are usually the first line of defense and a hacker's first roadblock to the control servers. They interact with the outside world on one side, and the operator's network on the other. They most likely transport both control messages and media. They are not only responsible for intercepting and discarding possibly threatening messages and media, but also to police the SLA (Service Level Agreement) which operators agree on among themselves, i.e. quantity of traffic and quality of service.

Examples: the IBCF and SEG are the two most typical forms of sentinel gateways in an NGN, protecting the network at the signaling level. The BGF on the other hand, performs the same function but at the media level. Sometimes these two functions are put together in a single network element and then we have a classic firewall functionality, or a session border controller (SBC)

Likely threats: sentinel gateways being the door to the network, should be the most resilient network element. Any DoS attack, whether directed at them or at the elements behind, will go through one of them. Unlike most of the elements behind them, their IP addresses will be global and readily known. It is also possible that since all signaling and media coming from a particular access domain will enter via a sentinel gateway, this element will be very attractive for redirection attacks, in which if an attacker gained access to the kernel, he or she could have access to media flows and redirect them to a receiving address for eavesdropping.

In this chapter, we have seen the types of attacks that network elements can suffer because of their nature and position in the network. In Chapter 7 we will examine the likely threats to an IMS network because of the characteristics of IMS, its architecture, its uses, and its users.

Now that the main types of NGN architectures have been introduced, and a possible generalization or abstraction of them has been proposed, it will be useful to investigate what research has been conducted on the security of these types of communications networks. We will do that in the next section, followed by an analysis of IMS security threats, and the introduction of security patterns.

PART B

SECURITY

6 CURRENT RESEARCH IN NGN SECURITY

There has not been any published work concerning applying security methods, measures, and tools to a “real” or “specific” deployment of IMS by a communications network operator. The components are there for this work to be done: 3GPP, TISPAN, IETF, CableLabs and other organizations have published extensively on the topic of security as it pertains to every domain and component of a real operator network. Products exist from multiple vendors to implement many of the technologies visited in this work. Methodologies such as Security Patterns exist and pattern catalogs have been published. But we are not aware of any work which could be used by an operator who wishes to deploy IMS, to bring it all together under a systematic process which evaluates threats, specifications, defenses, and methods and produces implementable recommendations.

The current literature does look at all of these topics in isolation and can be useful for delving into particular security questions. What follows is an extensive sample of recent research on security not only pertaining to NGN’s, but also to cellular communications, wireless networks, applications, and finally on security patterns.

The first set of papers is the most closely related to our present work:

In [Par09] Park et. al. evaluate the threats and vulnerabilities of a possible IMS deployment, and demonstrate possible successful attacks using the Georgia Tech OpenIMS core network testbed, which also contains equipment supplied by IMS core network vendors. Possible measures to mitigate the attacks are given, including full adherence to 3GPP prescribed standards. Other specific defenses against specific attacks are also analyzed in this work.

In [Qiu07] Qiu et. al. develop a model for the survivability of IMS at the control and application layers, based on Petri networks, but no study of actual security topics is undertaken. In [Hun07], Hunter et. al. explore general IMS security issues and the 3GPP specified measures designed to combat them, as well as other areas which could have an impact on security, like QoS, charging, enabling services, and regulatory considerations.

In [She06] – Sher presents the design of secure inter-operator communication by relying on the 3GPP NDS architecture and presents the mechanisms for establishing the IPSec security associations, Public Key Infrastructure, and Certification Authorities. [She09] explores the different attacks that an HTTP and SIP based IMS application server can be subjected to and proposes to defend against them by using TLS and Intrusion Detection Systems. One particular design for the IDS is presented and tested. Finally, in his PhD thesis [She07], Sher develops the most thorough work on IMS security that we have come across, giving a very good introduction to IMS, its architecture and vulnerabilities. He covers threats, air interface, key management, inter-domain security, authentication methods, generic bootstrap architecture, access network security, and other areas. He then proposes and develops a tool for an intrusion prevention and detection system for the IMS core and applications.

In [Sat09], Sathyan and Unni assess security risks to DRM media and its secure delivery over IMS networks. In [Kos07], Kostopoulos and Kufopavlou examine the security threats brought about by the heterogeneity of access technologies, service providers, and the need by roaming subscribers to access the IMS applications through these diverse accesses. The authentication mechanisms used in these cases are defined by TISPAN in the NASS subsystem. The NASS IMS Bundled Authentication is described, and measurements of the message sequence are taken using a simulated environment, and the results analyzed.

In [Pri09] Priselac and Mikuc look at problems with pre-IMS AKA access security, used with clients which still do not support IMS AKA, and rely on methods like HTTP Digest which need user password input. It’s noteworthy to mention that such mechanisms will not likely be implemented in any large network. They also look at insecure implementations of some security measures.

In [Mac09], the possible pitfalls with a mobile health application which bases its confidentiality and integrity assurances on a mobile operator-provided Generic Bootstrap Architecture (see Chapter 9) are examined, and an alternative is proposed. The authors argue that it may not be enough to trust the security of life sustaining equipment (e.g. a remotely controlled pump which delivers insulin to a patient) to a mobile operator or to the mobile terminal which contain the cryptographic key material.

[Bom02] provides good background information and thorough explanation of some of the topics introduced in this paper related to 3G mobile network security, such as: differences with respect to 2G security, access security, network domain security, and IMS security.

In [Bar06], Barkan, et. al expose critical flaws in the authentication protocol used in GSM networks, showing that cipher-text-only attacks are possible when the weak cipher A5/2 is used. They also show that protocol attacks are also possible even when the mobile terminal supports stronger ciphers, but also supports A5/2. Such an attack using a man-in-the-middle GSM base station is described. Their work was used to strengthen subsequent versions of the protocol and demonstrates the usefulness of publishing encryption algorithms for analysis by the general research community.

This next set of papers is more related to research in security patterns:

In [Kum09], Kumar and Fernandez introduce a pattern for a virtual private network.. In [Fer06], Fernandez and Pernul propose patterns for session-based access control, one of which is used in this work. In [Pel07], Pelaez introduces the concept of attack patterns, formalizing attack techniques so as to better understand how to defend against them.

In [Fer06a], Fernandez lays out the structure of security patterns, and surveys some existing patterns at the different layers where they apply (application, operating system, etc.). He then examines a methodology for applying patterns at all different stages of design, from domain analysis stage to implementation. In

[Fer07], Fernandez et. al. examine different VoIP architectures (H.323, SIP), possible attacks and possible security patterns.

7 SECURITY THREATS

Before we can investigate the security requirements of a specific NGN based on IMS, as we will do in chapter 10, it will be useful to categorize the types of threats that it will be exposed to. For the purposes of our work, it makes sense to do this by dividing the network into separate parts, i.e. the segment between two different network elements, or a particular layer, or the network component itself. Whereas in Chapter 5 we analyzed the threats based on the generalized network element type, here we study them based on the particular characteristics of an IMS network.

We re-use the diagram from the IMS architecture to show this in Figure 7.1

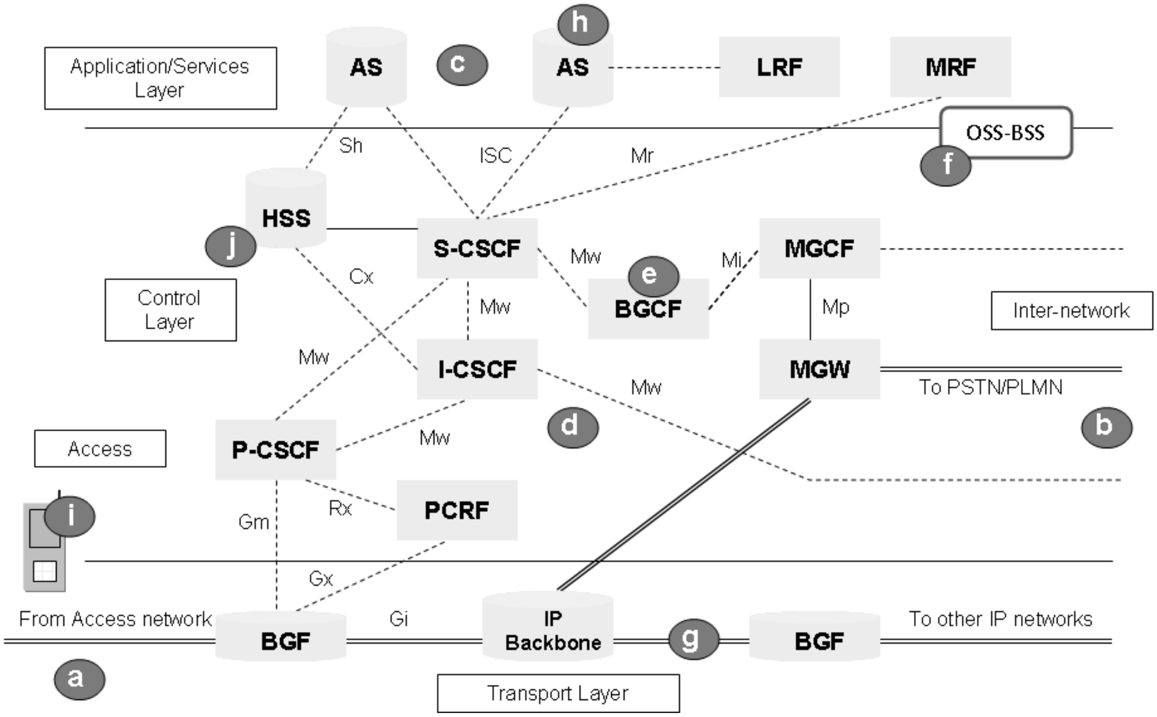


Figure 7-1 - Areas of Threat in an IMS Network

We have come up with 10 different areas to consider:

1. Access network (from User End-point to P-CSCF)
2. Inter-network (between operator A and operator B)
3. The application layer
4. The IMS core
5. The IMS network elements themselves
6. The Operations, Administration, and Maintenance plane (OA&M)
7. The media plane
8. Self admin web-services
9. Threats to IMS terminals
10. Physical access

7.1 Types of Attacks in the Access Network (from UE to P-CSCF)

The access network (AN) can be considered to consist of every link (physical or wireless) outside the P-CSCF, or if present, the SBC. Most operators will deploy SBC's in front of their P-CSCF's (or the decomposed version of the SBC comprising the BGF and BGC). The function of the SBC (type: sentinel) as described in chapter 5 and also in chapter 3, is to guard the protected domain from external attacks. In Figure 7.1, the access network will be everything to the left of the BGF.

Obviously, depending on the type of operator, there can be different types of access networks: a wireless provider will have an access network consisting of the radio access stations, some sort of concentrators/radio resource controllers, the backhaul links, and if the operator operates only an NGN, a packet backbone leading directly to the IMS; a fixed line operator will have either DSL lines, or IP-PBX lines, or direct IP connections, or Fiber to the Home, etc.; a cable operator will have the home cable modems, the coax facilities and other outside plant, most likely a fiber ring. The specifics of all this will vary depending on the technology, but it would be too much detail to include here. Also, no consideration is given in this paper to threats or disruption of the access facilities.

After a user has gained access to the access network layer 2 medium, in this case either an IP link connecting directly or indirectly to the BGF/SBC/P-CSCF, or a wireless channel, the types of threats to an IMS network are essentially not very different from those of concern to any type of computer network, with the exception that there is likely to be very little in the way of Windows operating systems within the IMS core itself, thereby rendering some attacks ineffective. The general attacks can be of the following four types:

DoS attacks - This type of attack tries to deny others access to the services provided by the IMS by severely busying out the resources in the network or the channels of communication to it. This is done by flooding the network with valid or invalid requests, generally from multiple sources.

Service theft - As the name implies, this attack seeks to gain unauthorized access to applications within the IMS or those facilitated by it, for example, VoIP, gaming apps, or premium video downloading.

Unauthorized access - This attack can be distinguished from the previous one in that whereas service theft only aims to obtain services for free, unauthorized access has a perhaps more malicious intent of causing damage to personal data or billing records, or inserting viruses or spyware for stealing information, blackmail, or even terrorism.

Network misuse - This type of attack is not destructive nor does it aim to steal resources, but is more prevalent than the others. One form of it is *spam*. In an IMS, spam is not only limited to nuisance e-mails or SMS's but can in principle be expanded to cover any type of multimedia communications, e.g. voice, video, music (imagine getting automated unsolicited video calls promoting pornographic sites). Spam can also propagate viruses as a secondary or even main intent.

Although other types of attacks are commonly mentioned in the literature, for example the Man-in-the-Middle attack, Network or User Spoofing, Eavesdropping, etc. these are just specific examples of, or ways to accomplish the generic attacks listed above.

7.2 Types of Attacks in Inter-network Communications

Unauthorized access – A network operator's biggest attack threat may come from other networks, since that is where the greatest potential base (in numbers) of sources of attacks is. Once another network's security has been compromised, the links to that network become paths for whatever has infiltrated the neighboring network. Viruses can try to traverse the demarcation line; hackers can seek out the internal topology, etc.

Service theft – Theft of services via the NNI interface can be of the same kind as that which can take place from the operator's own users, as described in section 7.1, or it can be of the service level agreement (SLA) type as mentioned above. An IMS operator will have SLA's with multiple IP network peers (IMS or not). These SLAs can be abused or misused through the neglect or errors in the peer network.

Dos attack – it is also possible to attack the network elements themselves via a peer network. This could take place against the first line of defense, usually the SBC, as in the access, or the attack could be against internal network element by trying to get a high number of valid service requests past the SBC.

7.3 Types of Threats from the Application Layer

Clandestine applications – As already described, applications in IMS reside at a different layer from the control core, and interface with the control core via two possible interfaces, one using SIP (the ISC interface), and the other using Diameter (the Sh interface). It should be the goal for an operator that there not be any communication between these two layers via any other mechanism. But even via these interfaces, internal topology information could be disseminated. If the IP address of the S-CSCF or the HSS becomes known to an outside entity, and those network elements are reachable, it would be possible for any SIP application server outside of the network operator's domain, or for that matter for any entity with IP connectivity to that NE to launch any of the attacks that apply to the user access listed in section 7.1. The attacks could be of a DoS nature at the IP layer or at the application (SIP) layer. This type of attack would not try to steal services or information from the IMS core but rather only disrupt its service capability. On the other hand, a clandestine SIP application server could try to mimic (spoof) an authorized SIP application server to obtain confidential information (user data, for example).

Attacks from legal (authorized) but compromised (i.e. manipulated to behave in a way other than its originally intended purpose) or “faulty” application servers, can also be considered. However, the remote likelihood of such an attack, and the difficulty of stopping it, necessitating a combination of very sophisticated SIP application layer gateways and intrusion detection systems, means that designing a defense for this attack would need to center on the physical and configuration control monitoring aspects, (i.e. protect the application servers themselves from being downloaded with harmful software) not on defenses inside the core.

7.4 Types of Threats from within the Core

Insider information theft – This would be the easiest type of security breach, however, this would be likely to happen only via the element management systems (EMS). It must be noted however, that many network elements also have local access ports. These must also be considered as being part of the EMS. The threats that arise via the EMS are examined in section 7.6.

Insider eavesdropping – The core elements in the IMS exchange sensitive data with each other about subscriber profiles, sessions, policy information, and billing. If all of this information is exchanged in the clear, and routed via switches and routers which are physically accessible to any maintenance personnel, the data is subject to being eavesdropped and used for purposes other than intended, with damaging consequences.

Insider malicious data destruction – This is similar to the first one. This is also only a risk via the EMS.

7.5 Threats to the IMS Network Elements

Malicious data destruction – If an attacker has been able to get through the external defenses and has gained access to an internal element, the intent is most likely to destroy data in the server or to install software which will harm the services offered.

Information theft – It is also likely that the purpose of the attack is to obtain information kept in the servers, in this case the Home Subscriber Server (HSS), particular to the IMS subscribers. This information is sensitive and its theft or changing could do great damage to the reputation of the network operator.

7.6 Threats in the EMS plane

It is clear that, with all the control that can be exercised via the EMS over multiple network elements in an IMS, the possible threats are practically unlimited. Physical access to an EMS may be easily gained by inside personnel who may not be authorized nor trained to manage the network. It is also a known fact that a large part of attacks to computing platforms or networks come from within the organization. The most common types of attacks that would be launched from the inside would be in the categories of malicious data destruction and information theft, which have already been described. Another imaginable attack would be simple sabotage, in which different network elements are re-configured, switched off, or programmed to do so at a certain time. This type of attack is easier to track than attacks that come from outside, with the greater probability of legal prosecution for the attacker. This makes them probably less likely than information theft.

7.7 Threats in the Media Plane

Eavesdropping - As with other types of communication, via electronic or other means, this is the most typical type of threat. Its purpose is to gain access to confidential information for financial gain or more nefarious intentions. An attacker intercepts communications between two or more peers while trying to remain undetected. The contents of the communication are not altered in any way as this would alert to his or her presence.

Integrity attacks – The purpose of this type of attack is to disrupt communications by rendering the contents in the media path unusable or by replacing some or all of the original contents with false information without the knowledge of the peers in the communications session. As with eavesdropping, this type of attack is most successful when undetected.

7.8 Threats from Self-Admin Web Services

Part of the appeal of many of the applications that can be delivered with an NGN is the fact that subscribers will be able to self-administer many of their features and options. The most typical example are the web portals where a telephony user can go to change the settings for features like call forwarding, voice mail, call blocking, viewing the detailed charges, etc. Depending on how the application is designed, some of these services may need a direct HTTP connection from the user's PC to the application server itself. This connection will not really traverse the IMS core network itself, but will possibly go through other NGN elements, especially routers (which may also route real IMS signaling and media). This easy HTTP access via critical network elements and to the IMS application server itself presents perhaps many more opportunities for traditional Internet attack methods than all other threats combined.

It will be very important to design the necessary safeguards into this type of network access.

7.9 Threats to IMS Terminals

An IMS terminal is any computing device with a SIP client (a SIP client is just a software application which complies with IETF RFC 3261 [ROS02]). IMS terminals may come in all kinds of form factors and hardware platforms: a lap top, a flip phone, a PDA (Personal Digital Assistant), an adapter box, a DSL router, a DOCSIS modem for cable TV networks, a satellite phone, a TV set top application, a refrigerator or air conditioner control thermostat (some day). Depending on the operator and the terminal vendor, there may be additional requirements or design on top of the SIP client, to allow it to use the features for which it was meant. All these devices have one thing in common: they contain software which is allowed, with the proper credentials, to access the IMS network and some of its applications.

It is clear that all of these “appliances” can be attacked and can contract the same type of viruses that threat personal computers. It is not the object of this work to solve that problem, only to see that if the above happens, the operator’s network is as immune as possible to threats coming from these now penetrated IMS terminals.

7.10 Physical Access

Lastly, it is clear that any type of infrastructure has the potential of being physically attacked. The benefits of protecting computing elements, links, power sources, etc. from physical access have to be measured against the need to have legitimate access to such equipment for operations, maintenance, and even cleaning personnel. There must be a balance between both of these requirements.

We have now built a fairly significant body of data related to how a Next Generation Network can be attacked. In Chapter 5 we analyzed this from the point of view of the network elements themselves, and their characteristics and placement within the network, whereas in this chapter we have looked at it from the point of view of the specific IMS architecture and its use by operators and end-users.

We are now ready to look at the other side of the problem, namely, the tools at our disposal to defend against these attacks. In Chapter 8, we look at a body of knowledge called Security Patterns and introduce some new ones. In Chapter 9 we investigate specific IMS security tools derived from the patterns in Chapter 8, and from existing patterns in the literature.

8 SECURITY PATTERNS

Patterns are an analysis and design methodology that defines a vocabulary that concisely expresses requirements for a particular system, without getting into implementation details [FER08]. The description of architectures using patterns makes them easier to understand, provide guidelines for design, provide a possible simulation model, and allows designers to compare and match methods or building blocks to their needs. Security patterns in particular [FER06a], apply this methodology to the analysis of security requirements for infrastructures, networks, and other designed systems. Security patterns for networks aim to study the possible ways that a computer network can be attacked and to provide the high level design tools to defend against those attacks. Conversely, misuse patterns have also been defined and studied [FER06b] with the purpose of finding weaknesses by looking at the problem from the other side. A considerable number of security patterns have been developed and some of them are referenced in Chapter 9.

Having analyzed the possible threats to NGN's in the previous chapter, we now focus on describing 6 additional patterns to those already available in the literature, which are more related to the security needs in an NGN. Together with already existing patterns which are introduced in Chapter 9, we will have a tool set of 13 patterns, with which to undertake the security design for the case study. For the time being our intention is just to fill certain deficiencies in the existing set of patterns with the 5 new ones introduced here. In Chapter 10, where we will finally undertake the task of designing the security architecture, we will show how each pattern (if in fact all of them are used) responds to the identified threats.

The patterns introduced in this chapter are the following:

a) Separation of Functions

b) Network Element Hardening

c) Restricted Network Layer Communications

d) Topology and Information Hiding

e) Precise Application Layer Communications

f) Automatic Deregistration

We now proceed to describe them

8.1 Separation of Functions in Communications Networks Pattern

Intent

To simplify the network architecture by separating the communications network into layers or domains, grouping within each elements with similar functionality. Unlike in Chapter 5, the focus here is on the network as a whole and not on the network elements themselves.

Context

All new Next Generation Networks deployed by telecom operators are IP-based. These networks are being deployed not only to offer traditional fixed or cellular voice connectivity, but also to handle many new types of services such as: video calling, messaging, push-to-talk, conferencing, multimedia messaging, see-what-I-see, video conferencing, and presence-based services to name a few.

Problem

Unlike the TDM-based networks that were deployed up to now, where the intelligence for routing, call control, and applications was centralized in one network element (the Digital Exchange or Mobile Switching Center), IP-based communications networks are comprised of many equally complex network nodes, each possibly responsible for completely different functions. This decentralizing of intelligence diffuses the security demarcation lines, opening up new points of attack. If the function of each network element and its place in the network are not clearly delimited and understood, it will be difficult to analyze it in terms of its security requirements and to put in place the defenses needed.

The solution to this problem is affected by the following forces:

- Non-arbitrary classification. The different network elements need to be classified into logical categories based on the function they perform; otherwise, it is difficult to make them secure.
- Limited number of categories. There cannot be a large number of types of functions, otherwise the end result will not help in achieving clarity.
- Ambiguity. There may be certain network elements which perform functions that might categorize them in two separate areas.

Solution

Define different domains based on the function of the network element and partly on whether it carries media packets. Interface these domains to each other by well defined protocols. The functions of the network elements (NE) can be classified as shown in Figure 8.1:

- Transport – the NE's main function is to transport media packets between two of its interfaces. It receives control information from NE's in the Control domain (see below) which it uses to select its input and output interfaces, and to apply or not other media-affecting functions such as transcoding, gating, rate-limiting, etc.
- Control – the main functions of such NE's are to accept session requests, set up sessions, steer them to the right end-point, whether this be application servers or other end-users, keep records for charging and for other purposes. Control NE's should not see any user-to-user or user-to-application media flows. In some circumstances, control NE's may originate media towards the user, as in the case of an announcement or tone server, to provide the session originator call progress indicators.

- Application – NE’s in this domain provide the real services for which the end-users are willing to pay: voice communications, voice mail, conferencing, video streaming, push-to-talk, multimedia messaging, etc. Session requests get steered towards the application domain by the control NE’s.

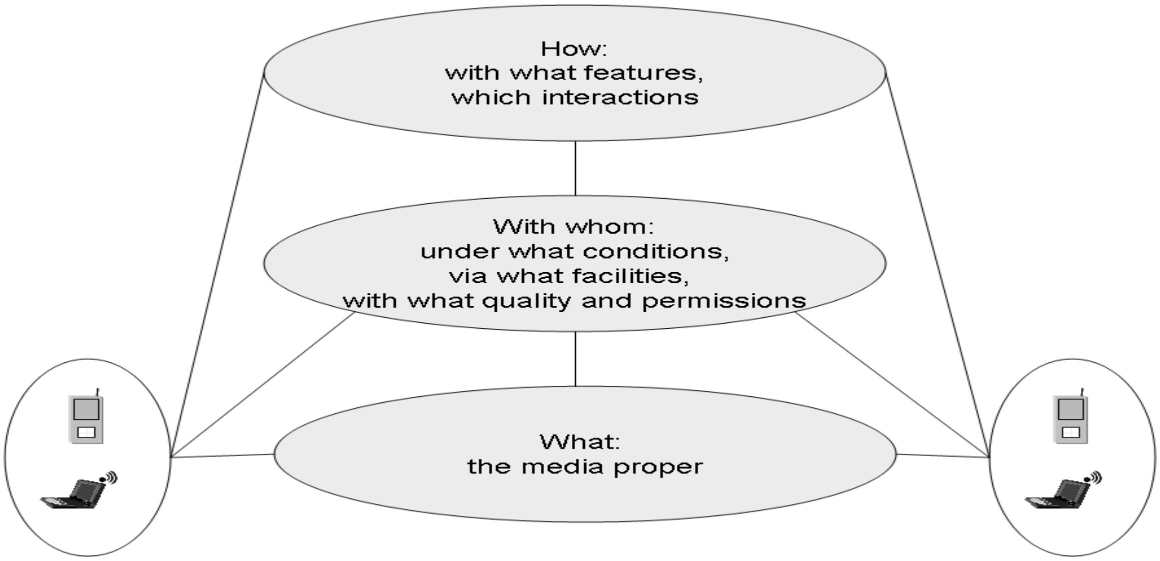


Figure 8-1 – Separation of Functions

Consequences

It is easier for the security architect to determine types of attacks possible within each of the three layers, and therefore easier to decide what defenses to use. The classification introduced was non-arbitrary in the sense that the three layers are derived naturally from the purpose of a communications network. For the same reason, there are also no more layers than needed. Given this, it will be possible to compartmentalize the security analysis. At this level the ambiguity is not evident nor important. It is only when we try to use the pattern to realize a communications network that decisions will have to be made as to how to classify particular elements.

Known Uses

The 3GPP architecture makes use of this pattern as does TISPAN. Before them, the Multiservice Switching Forum (MSF) architecture also separated media from control, but it did not separate control from applications. To some extent, the PacketCable (pre 2.0) standards also used a similar but 2-tier separation.

Related Patterns

In [BUS96] a Layers pattern is presented which provides another way of separating Concerns. This pattern is a special case of a more general principle of Separation of Concerns

8.2 Network Element Hardening Pattern

Intent

To make each individual network element more resistant to software attacks by reducing the target space.

Context

The telecom and multimedia networks addressed here are unlike previous generations of communications networks in that while the former were for the most part based on proprietary hardware and operating systems, the current ones are based on commercial off-the-shelf hardware (COTS) and operating system (OS). Even the programming language that was used then, was sometimes special purpose (e.g. CHILL, which stands for CCITT High Level Language). Today, most communications software is developed in C, C++, or JAVA.

Problem

Network Elements and application servers running on COTS platforms, programmed using an environment (e.g. Solaris, Linux) and a language commonly known by thousands of programmers (e.g. C, JAVA), are much more vulnerable to attacks than their predecessors with proprietary environment, not only due to the widespread use of the operating systems, but also to the fact that the COTS platforms sometimes come pre-loaded with additional software necessary for common IT uses. This is exacerbated by the use of IP networks to connect these NE’s and AS’s, since widespread familiarity with that protocol (suites of protocols), also makes it easier for potential attackers to get to the target NE’s.

The solution to this problem is affected by the following forces:

- Effort. This is twofold: first there is the effort to ascertain the vulnerabilities of each individual platform (NE or AS), to asses which of the many packages loaded in the course of fitting it with the needed environment (OS) are really necessary and which are superfluous. Second, the continuing task of maintaining a team which is up to date on the latest security risks found by the software community (e.g. Carnegie Mellon Software Institute), and to bring these changes into the platforms.
- Risk. This is the possibility that one or more of the hardening actions renders inoperable a feature that is necessary now or in the future.
- Coverage. No amount of hardening will result in complete security; we are satisfied with controlling the most common types of attacks.

Solution

In order to make a platform based on widely available and well known components harder to attack, all interfaces and functions which are not needed for the intended use of the network element must be closed and disabled.. This process, represented in Figure 8.2, has come to be known as platform "hardening" in the industry. Hardening takes place during the design and development process, and should be tested throughout up until and including deployment. Unlike other security methodologies, where the solution lies in adding something, the intent here is the opposite in that it consists of taking out as much as possible while leaving enough of a kernel for the application to be able to do its job.

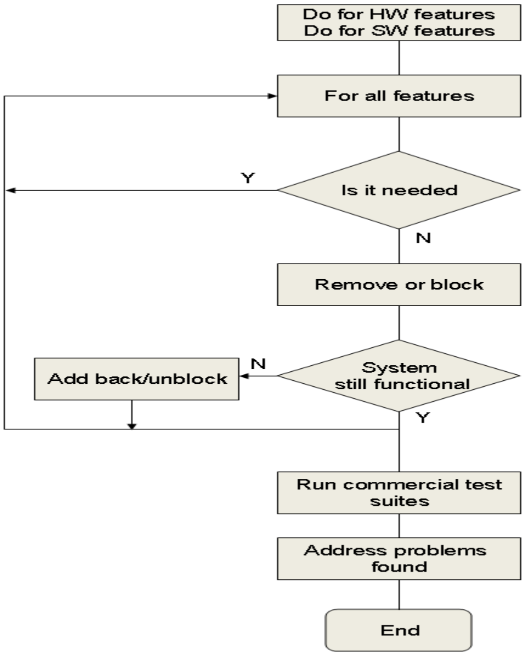


Figure 8-2 – Platform Hardening Process

Some examples of hardening are:

- Removing unnecessary physical accesses to the machine (e.g. Network Interface Cards)
- Disabling ports in partly used NIC's.
- Disabling unused sockets (IP address-port combinations)
- Disabling or removing unused OS API's
- Turning off OS functionality which is not needed by the application
- Removing pre-loaded application software which is not needed by the intended application.
- Iteratively running commercial vulnerability suites against application

Consequences

The platform is stealthier on the one hand, because it contains fewer doors through which to perpetrate attacks and because it will not respond to discovery queries from attackers intended to divulge its presence and capabilities. It is also less prone to failure on its own, since many functions are rendered inoperable, which means less code is running. It will have also gone through more rigorous testing in the process of being hardened. The hardening of the platform, and its coverage, are a tradeoff between effort and security. It is clear that a benefit/cost ration analysis needs to be performed before undertaking this and an appropriate balance should be reached. The risk factor is more controllable. During the process of hardening a platform, it will become clear what features are rendered inoperable and whether they are or will be needed for legitimate use.

Known Uses

Most servers used today in critical fields such as telecommunications, military, and banking, undergo hardening if the platform is commercial off-the-shelf hardware and operating systems.

8.3 Restricted Network Layer Communications

Intent

To limit the potential sources of attacks to individual network elements in an NGN by restricting its communications partners.

Context

One of the characteristics of the new IP-based voice and multimedia communications networks, is the flexibility to add applications (usually by installing completely new servers), to increase the capacity of the network by growing horizontally (more servers), and to even enhance the worth of the network by altering its architecture and adding more functions as the standards progress. An example of the last one is the addition in 3GPP release 7 of the new function E-CSCF, on top of the already existing P-, I-, and S-CSCF. The E-CSCF (for Emergency), formalizes the procedures to provide emergency call services in IMS networks.

Problem

IP communication networks are basically “open” networks. Their manufacturers even use the term “open” to indicate inter-operability with the components of different vendors. Not only is the IP suite of protocols open for anyone to use, but the higher level protocols developed on top of IP, such as SIP, SOAP, XML, Radius, Diameter, RTP, etc. are also defined (usually) in IETF RFC's and standardized to the latest detail in the bodies which define the architectures discussed here (3GPP, 3GPP2, TISPAN, PacketCable). This openness, gives the opportunity to possible attackers who have managed to breach any outer defenses, to spoof legitimate communications partners (other intra-network NE's), and thereby compromise security.

The solution to this problem is affected by the following forces.

- Auto discovery. To preserve the flexibility of the IMS architecture, it must be possible to introduce new control elements or application servers; such that they are recognized immediately as legitimate communications partners, without having to manually, or via great effort, modify the database in all existing NE's. Likewise, the chosen solution must be able to easily cope with any changes to the topology of the network.
- The specific communications partners of a node may dynamically change, and network elements need to be able to adapt to change in a convenient way.

Solution

Restrict the communications partners of each NE by predefining with which other nodes they can interact. All network elements must be provided with information, either statically, or dynamically, via some intelligent identification mechanism, on which other nodes in the network are allowed to be their communications partners. This may need to be done at more than one level. For example, NE1 must know that it can only communicate with NE's in the network having IP addresses: IP1, IP5, and IP9, or, it may additionally be told that it, being an S-CSCF, can only logically communicate with other S-CSCF's, with all I-CSCF's, with all P-CSCF's, with the HSS, with the BGCF, with application servers AS1, and AS2, and with DNS servers A and B, for example. Such knowledge in the NE, is commonly referred to as a "white list". A white list is a list of those elements with which an NE can communicate.

Any messages from a source that is not in the white list get rejected. As described above, the white list in an IMS network may contain physical identifiers (i.e. IP address), functional identifiers (i.e. "what function do I play in the network), or even logical identities (i.e. “not only am I an I-CSCF, but I am this particular I-CSCF”).

Consequences

Communications within the operator’s own IP Voice and Multimedia network NE’s is safer, as well as with NE’s in peers’ networks. However, administration effort increases in order to populate the white lists with allowed partner identifiers. The problems of auto discovery and dynamic configuration change are avoided if the white lists are populated with Fully Qualified Domain Names (FQDN) instead of absolute IP addresses. If this is done, additional network elements of the same function can be deployed by just adding them into the DNS server, which load balances over multiple instances of the same function.

Known Uses

Closed private networks, or open networks using firewalls, which allow messages and/or media to come from certain IP addresses or IP ranges. Full cone NAT devices, which create a white list on demand, based on what peers the user behind the NAT chooses to communicate with.

Related Patterns

This is a special case of the Firewall pattern, which can establish other parameters in addition to white list as condition to permit communication or not.

8.4 Network Topology Hiding

Intent

To limit the amount of sensitive topology information, which leaves the secure domain towards the peering networks and the subscriber access. Limiting this information reduces chances of attack.

Context

Unlike in previous generations of communications networks, with the exception of the SS7 network, where there was no “routing” as we know it now of messages via intermediate nodes, an IP message is routed based solely on the destination address. Any NE can be addressed by an attacker if its IP address is known. There is no need to have direct physical connectivity to it. So for the same reason that a person might want to keep her address, phone number, or e-mail address confidential, to prevent unwanted communications, a network element in an IMS network is safest when its address is not known outside its domain.

Problem

When the addresses of network elements, or more generally, when the topology of a network is known to elements outside of the network, one of the key safeguards of security, anonymity towards third parties, has been surrendered. To a large extent, a hacker will not attack that which he or she does not know it exists. If the address, or the FQDN of the HSS, for example, the Home Subscriber Server, which contains the entire subscriber database of an IMS network is not known to any NE’s outside the operator’s domain, the subscriber information is that much more secure.

The problem is that many protocols, SIP among them, contain vital information about the sender in the IP headers as well as in the application layer fields.

The solution to this problem is affected by the following forces:

- There are cases in which there is a legitimate need for information about the source or destination of the message, either at the IP, transport, or application layers. In those cases, the information cannot be suppressed.
- Hiding this information using encryption may require decryption in other nodes, also possibly transient ones.

Solution

When exchanging information with other networks or with end users, content which may reveal internal topology or other network sensitive information, should be suppressed before sending the message. If that information is needed by a network element beyond the network edge, then the content cannot be suppressed, but at least it must be encoded.

A practical way of achieving the above is by using a Topology Hiding Inter-network Gateway (THIG), see Figure 8.3. All messages leaving the operator network are routed via this network element, which performs the actions mentioned above. At the same time, this interface into the network is all that the partner operator sees.

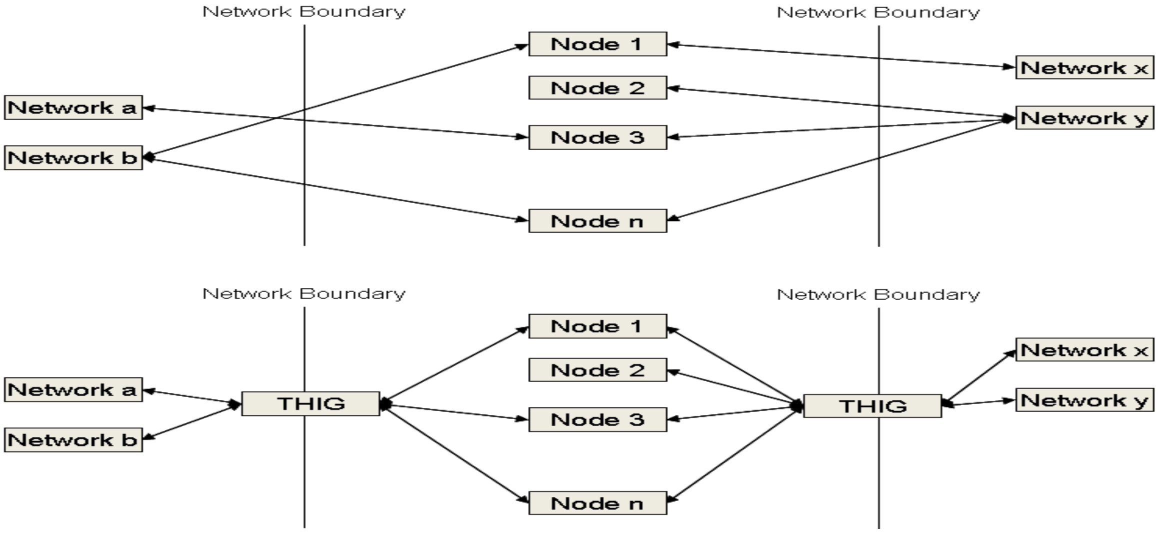


Figure 8-3 – Network without and with Topology Hiding

Consequences

Topology hiding keeps potentially sensitive operator network information from leaving the trusted domain. Attackers are not able to collect this information from messages leaving the operator's domain, to try to find out what functions the different NE's perform, or to launch blind attacks against particular addresses.

The THIG can be provisioned to suppress information depending on the needs of the partner networks and of the applications going through, so that no essential information is witheld. With this particular solution, no encryption is necessary.

Known Uses

Information hiding is a common practice in structured and object oriented programming. When that information is network topology, firewalls, session border controllers, and to a certain extent media gateways, also provide that functionality. All major next generation networks standards bodies, 3GPP, TISPAN, PacketCable, have standardized topology hiding in their specifications.

Related Pattern

Information hiding in computer programming.

8.5 Strict Application Layer Communications

Intent

To decrease the chance of application layer attacks by not accepting any messages that deviate in the least from the corresponding standard.

Contex

There are only a handful of protocols used in most reference points in 3GPP-based networks at the application level. Among those, SIP is responsible for most of the control and session routing, between the end-user device and the control layer, the control layer and the application layer, and the control layer and peer networks. All three reference points lay (possibly) outside the complete influence of the network operator. Another protocol for which this can apply and which is also critical in the operation of the network is Diameter. It is key because it’s used for subscriber database access by control elements and applications, and as such, when the applications are outside of the operator’s control, this reference point can also be outside of the total influence of the operator.

There are other protocols which could also be covered by this pattern, but are not seen as vulnerable to the problem, or are not as critical to the operation of the network as SIP and Diameter.

Problem

SIP and to a lesser extent Diameter are loosely defined protocols. SIP in particular, is very much a living protocol in that new extensions are often proposed by means of new IETF drafts, which then may be promoted to RFC's. In addition, organizations that use the basic SIP RFC's as the basis for the network architecture, like 3GPP, often modify or enhance the basic set of RFC's by adding new parameters in their own Technical Specifications. This constant enhancing of the protocol coupled with its openness is one of the benefits that proponents of next generation voice networks claim: unlimited possibility of applications and flexibility. But this is also what can turn into severe vulnerabilities for the same networks.

The solution to this problem is affected by the following forces:

- Dictates of the protocol itself. The Robustness Principle of the Internet, first stated in an RFC in RFC 760 [POS80], says: "In general, an implementation should be conservative in its sending behavior, and liberal in its receiving behavior. That is, it should be careful to send well-formed datagrams, but should accept any datagram that it can interpret (e.g., not object to technical errors where the meaning is still clear)."

- Interoperability and future extensions of the functionality. We need to balance out the need for promoting interoperability and extensibility.

Solution

Those network elements using SIP and Diameter protocols and communicating with entities outside the secured domain (dotted line in Figure 5.1) should have an administrable option to set their protocol behavior to reject by default any messages with undefined fields, instead of just ignoring them and continuing the message processing. If the message is processed further, and allowed to make it on to higher layers or to other elements, possibly harmful content could be allowed in. By stopping the analysis at the lowest possible point, risk is minimized.

Consequences

Messages which are structurally sound but which contain at least one undefined or not understood parameter will be rejected, making it more difficult for a hacker to launch successful protocol attacks. In NGN for communications, the robustness principle is not as critical as in the Internet; the reason is the so called IOT's (Interoperability Tests, or events) which evey operator performs with all the vendors of its NGN, including end-devices, to make sure that interoperability is achieved. Any problems found due to the implementation of this pattern can be solved before commercial launch.

Known Uses

In legacy telecommunications networks, protocols such as Signaling System number 7 (SS7), adhere to the standards much more closely than text-based protocols such as SIP or HTTP. In data communications networks as well, with X.25 being an example. The 3GPP has also defined rules in its latest release for a Topology Hiding Inter-network Gateway.

8.6 Automatic De-registration

Intent

To minimize the chance of attacks via valid user end-points by agents other than the legitimate subscribers by taking away privileges after a long time of inactivity.

Context

One critical difference between voice and multimedia over IP and previous generation communications networks is the concept of registration. In regular communications networks, a device is associated with a physical location, i.e. the port number at the central office's digital line unit. By extension, the person or persons that can be reached via that device are reachable at that same location. Cellular telephony removed one of the restrictions, physical location, but it still retains the relationship between device (cell phone, or in the case of GSM, the SIM card) and the person associated with the corresponding phone number. With IP telephony based on SIP, a user can register at any compatible device anywhere in the world and use that location as the place where calls should be received.

Problem

The flexibility offered by SIP telephony to the end user, to be able to register from any SIP device via any type of access network and to be able to receive voice or multimedia sessions anywhere in the world, brings with it new vulnerabilities. When a user registers with the network, by sending a REGISTER message to the registrar and providing it with his or her authentication vector, anyone with access to that device will have access to the network and will be able to initiate a number of attacks against it. This could happen, for example, if a user leaves a laptop connected in a hotel room, after having registered. It could happen in a corporate network in the same manner. It could even happen at a friends home if a user has temporarily registered via the friend’s home computer to receive calls there.

The solution to this problem is affected by the following forces.

- Registration is a prerequisite for service. If registration is made too difficult, the quality of service perception will be poor.
- Registration which has to be performed too often will drive up usage of network resources (links, processors), negatively affecting performance.

Solution

For security purposes, a network will accept a registration from an end user, as a time-limited contract to deliver subscribed services. After the time limit, which will be administrable by the operator, expires, the contract to deliver services expires and the end user will need to re-register with the network. When a subscriber is not registered with the network, no messages (other than a valid registration request and emergency calls) are accepted.

Consequences

The capacity for attacks to come from devices or peers with limited registration periods is greatly reduced. Standards and IOT's can be used to strike a balance between the effectiveness of the security measure and the needs of the applications running over the network, so that quality of service perception and usage of network resources remain acceptable.

Known Uses

"Limited registration contracts" is a feature already implemented in client-server computer networks, web applications and in other applications like some remote wireless e-mail (Blackberry or Intellisync from Nokia).

Related Patterns

Most IT systems and personal computer operating systems implement a similar measure for login sessions. In this chapter we have introduced 6 new patterns to use in the security architecture of an NGN. In the next chapter we will see how other security patterns, already defined elsewhere, can also be used for our case study. We will also look at specific examples pertaining to IMS.

9 SPECIFIC IMS SECURITY TOOLS AND METHODS

In this chapter we will review the specific cases, derived from the patterns described in chapter 8 and additional ones previously published elsewhere, which have been used, or can be used, for the security of IMS networks, and which we will apply in chapter 10 of this work.

From each of the patterns in chapter 8, and from other patterns defined elsewhere, we can select one or more "tools" as necessary:

a) Separation of Functions Pattern
b) NE Hardening Pattern
c) NE Communications White List Pattern
d) Topology Hiding Pattern
e) Strict Protocol Adherence Pattern
f) Time Limited Registration Pattern

In addition, we can select the following additional patterns, which are part of the existing body work in security patterns, and which also apply to IMS:

g) Authentication Pattern
h) Authorization Pattern
i) Communication Confidentiality Pattern
j) Secure Channel Pattern
k) Non-re pudiation Pattern
l) Intrusion Prevention Pattern

m) Intrusion Detection Pattern

n) Demilitarized Zone (DMZ) Pattern

o) Virtual Private Network (VPN) Pattern

p) Security Policy.

We now proceed to describe which known protocols, tools, procedures, or best practices exist today which we will be able to apply to our two network examples

9.1 Separation of Functions Pattern

1) Clear demarcation between the Application, Control, and Transport layers

This is an inherent property of the architecture selected for the two networks, the IMS. However, since the 3GPP IMS standards do not enforce this, but only recommend it as a desirable design goal, an operator implementing an IMS-based network may be able to bypass this recommendation, since there will always be products available which do not conform. We will be conscious of this security pattern when designing the security architecture of our two operators. For example, the following best practices will be adhered to:

- All the core functions, I-CSCF, P-CSCF, S-CSCF, etc. will reside on different hardware platforms and communicate with each other only via the defined interfaces (i.e. no internal proprietary protocols)
- There will be a clear demarcation between control elements, databases, and applications

9.2 NE Hardening Pattern

This pattern has specific examples in both the use of certain tools, and the adherence to known best practices. Below are some in each category which will be used:

1) Application of operating system provider recommended hardening techniques (Solaris, Linux, proprietary OS, Windows if used)
2) Removal of unused services
3) Follow guidelines for secure OEM hardware and software configuration
    a. DNS
    b. Routers and switches
    c. Firewalls
4) Perform all manufacturer recommended security scans and tests using available commercial tools below and/or others
    a. Nessus [NES00]
    b. Codenomicon [COD00]
    c. SiVus [SIV00]
    d. Specific tools for testing the resiliency of the network to SIP and RTP flooding attacks
    e. Specific tools aimed at testing resiliency against malformed messages

9.3 NE Communications White List Pattern

This is not a tool or an algorithm, but rather a “best practice”. But in order for the operator to be able to put it into practice, it must be ascertained during the procurement phase, that the vendors of the IMS equipment provide this functionality in each of their network elements. During the network implementation phase, the operator, together with the system integrator, will need to define these white lists, implement them in every NE, and test them. Examples of particular implementations are:

1) The network HSS can only receive requests from the EMS (Element Management System), I-CSCF, S-CSCF, and any application servers in the network which use the HSS as their data repository. Messages from any other IP address are rejected at the lowest possible level, and logged or alarmed at the operator’s option.

2) The BGF (Border Gateway Function) can only receive messages from other peer network BGF's, as defined by the operator. All other control messages or RTP flows will be discarded and alarmed.
3) The EMS system cannot receive any messages from any NE's other than those it controls. Communication with other remote workstations must be via VPN.

9.4 Topology Hiding Pattern

Topology hiding concerns also the correct application of features in the protocols and in the border elements, which must also be provided by the IMS equipment vendors in their products. It will be important in this respect to follow any recommendations in 3GPP TR 24.229 [3GP229] concerning this area. Topology hiding is also a consideration when designing the overall architecture, in that for example, an additional routing network element can be inserted as proxy in between the operator's domain and any other network. These three possible applications of this pattern will be taken into account in the design.

9.5 Strict Protocol Adherence Pattern

In practice, there is very little that the operator can do to implement this pattern during the network implementation phase. This is more of a consideration during the requirement specification phase. It is then that an operator can inquire from all vendors what their design philosophy is with respect to protocol handlers, and can request additional safeguards against non standard messages, as well as other features like logging and alarming. In our two examples in this paper, it will be assumed that the operator has done its due diligence during the specification phase.

9.6 Time-limited Registration Pattern

This pattern will have several applications in our example NGN networks. The first one will have to do with the basic feature of SIP which requires that all users in the network register with the network registrar, usually the S-CSCF and the HSS, before any services can be provided. The second one applies at the administration level, where all operations personnel must also log in (register) into the EMS before any administration work can be done. In both instances, a registration must not be valid longer than a predefined length of time, administrable by the operator. In the case of SIP end-users, the registration will expire unless a REGISTER message is received periodically from the end device (e.g. 60 minutes). In the case of the operations workstation, a user is automatically logged off after the pre-defined inactivity period.

9.7 Authentication Pattern

The authentication pattern is also applicable to the same two areas as the previous. Subscribers must authenticate themselves to the network (and the network must authenticate itself to the subscribers), and operations personnel must be authenticated by the network before being allowed access. In addition, some services at the application layer, for example third party applications which are accessible to subscribers of the operator's network may also require authentication. Authentication is also a requirement for access to the EMS from remote workstations via VPN's.

In the case of subscribers, there are several protocols and algorithms which have been standardized by the 3GPP and can be used in our examples:

a) Early IMS Authentication [3GP978]

b) IMS AKA Authentication [3GP203]

c) Digest Authentication [ROS02]

d) UMTS AKA Authentication for subscriber mobile packet access to UMTS networks [3GP102]

For centralized authentication by third party applications the following standards can be applied:

e) Generic Bootstrapping Architecture (GBA) as defined by 3GPP in TS 33.220 [3GP220]

f) Liberty Alliance for web-based Single Sign On [LIB00]

g) Identity Federation as defined by 3GPP in TS 33.980 [3GP980]

For operator access to the IMS element management systems, best practices with respect to implementation of administration ID's for all services accessible via the network and element management systems, will have to be adhered to.

9.8 Authorization Pattern

The authorization pattern is closely associated with the authentication pattern. After a subscriber or operations personnel are authenticated and therefore granted access to the network services, each request for one of the services will have to be authorized. The functions to conduct the authorization will have to be already integrated into the different network elements. It is again a matter of ascertaining during the requirements phase that these software tools are in place, and of following best practices (and having a policy in place that describes and monitors these practices) during the implementation phase. The functions that will need to be available in the software to be able to authorize subscribers and users will include algorithms in the category of Role-Based Access Control and Reference Monitor patterns [FER08a]. We would expect these in any network elements that make decisions about how to handle end-user requests for services (telephony, messaging, etc.), like the S-CSCF, and in those elements involved in network and element management.

9.9 Data Confidentiality Pattern

This pattern is applicable to data which must be protected from theft even when not being transmitted across an interface to another network element. An example is passwords and other subscriber or operator data which can be displayed from an element manager system terminal

9.10 Secure Channel Pattern

This pattern is applicable to multiple physical segments of an NGN:

a) subscriber media confidentiality (voice, texts, video)

b) subscriber-to-network signaling confidentiality

c) intra-network element signaling confidentiality

d) inter-network signaling confidentiality

In every case, confidentiality involves protecting the contents of the communication, whether data or signaling from eavesdropping by any unauthorized parties, for any reason. In the case of Law Enforcement Intercept (LEI) as described in chapter 2, confidentiality can be violated following a legally issued court order.

Confidentiality protection usually means encryption. In the IMS specifications there are two important algorithms specified for encryption of signaling and data: the IPSec and TLS protocols.

Recommendation TS 33.203 [3GP203] specifies the IMS AKA algorithm, which is not only used for authentication, but also for agreeing on the keys (KA: key agreement) to be used by IPSec for the protection of subsequent communications. Recommendation TS 33.310 [3GP310] and IETF RFC 3261 [ROS02] specify the framework for applying TLS.

Recommendation TS 33.210 [3GP210] specifies the framework for implementing intra and inter-network security with IPSec. It defines reference points Za and Zb as shown in Figure 9.1

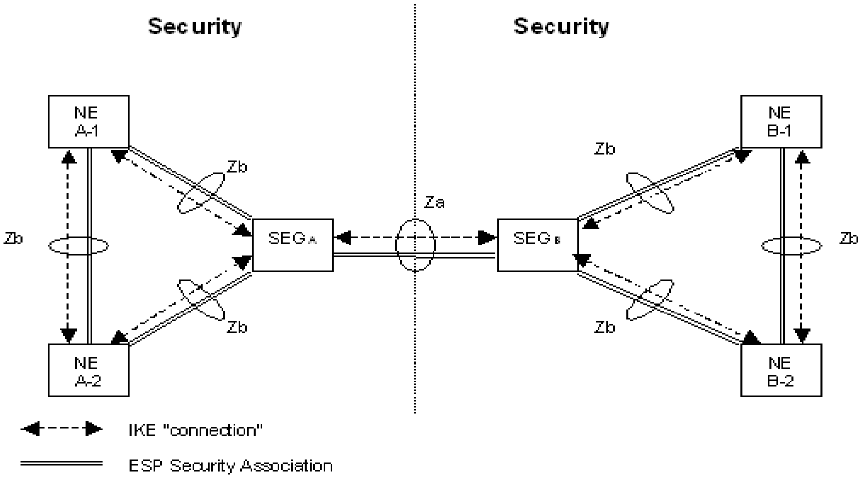


Figure 9-1 - Abstracted NGN Architecture

9.11 Communication Integrity Pattern

In its strict definition, communication integrity is a very basic part of any NGN protocol. There are sufficient mechanisms in place to guarantee that messages are not changed either deliberately or due to transmission error. There are attacks that can be perpetrated however, like the man-in-the-middle attack (MITM), which will go undetected by basic transmission integrity mechanisms, because they operate at a higher level, i.e. by impersonating the rightful communications partner.

The first class of integrity detection functions will have to be in place at every interface and this is something that is ascertained during the requirements phase. For MITM attacks, there will be other tools or features within the protocols themselves which will need to be exercised. We'll see how in chapter 10.

9.12 Non-repudiation Pattern

Non repudiation is not a classic security problem in telecommunications networks. Nevertheless, since the SIP protocol used by IMS requires subscriber authentication before any services can be provided, the only time in which this repudiation can be tried is when access to a device is compromised. This would be similar to someone gaining access to a person’s e-mail account, and using that access to send malicious e-mails.

9.13 Intrusion Prevention Pattern

Intrusion prevention aims to detect the presence of malware (malicious software) in downloaded files, and to warn the user. If a file with malware is detected, the user usually is given the option of deleting or isolating the file, in order to prevent infection of the host system.

Intrusion prevention software, which is available both commercially (preferred) and as open source, should be installed in those machines, like network and element management systems, where access to the internet or to vendors remote management systems are allowed.

9.14 Intrusion Detection Pattern

There are commercial as well as open source applications of the intrusion detection pattern available, for example SNORT® or Sourcefire [SNO00], which can be deployed behind the IMS session border controllers and/or firewalls, in order to alert the operator when these devices may have been breached and an attack is possibly underway. For each of the two NGN examples presented, we will have to decide where IDS should be deployed (basically, in just some subnets or VLAN’s, or in the entire trusted domain).

9.15 Demilitarized Zone (DMZ) Pattern

This pattern does not consist of a special protocol or tool which must be used, but is rather a design construct by which network elements which can be access from (or have access to) both the public and the secure domains in a network are placed between two firewalls, one towards the unsecure domain, and another towards the secure domain. Appropriate access control lists (ACL) and filters are put in place in the internal firewall to limit access to the secure domain to messages coming only from the elements within the

DMZ. The outside firewall will similarly contain a set of rules of what type of access is allowed, permitting only a certain number of protocols, to a certain set of IP address/Port combinations, etc. Among the rules that can be built into a DMZ is also those that permit only session requests (for any protocol or higher layer application) in a given direction, while rejecting requests in the opposite direction.

9.16 Virtual Private Network (VPN) Pattern

This pattern could be considered a combination of the White List Communications and the Communications Confidentiality patterns. Its goal is to define a “channel” for communicating with only a single entity, and to do so with data encryption. This channel is not limited to single hops, but can traverse any number of routers in between the end points. Typically, the encryption used is set up by using IPSec.

9.17 Security Policies

Finally, it is important that all the security mechanisms, tools, and practices be organized and put in place as part of an overall published security policy, known and available at all relevant levels within the operator. The implementation of this policy should be under the responsibility of one single security director, who shall regularly monitor its compliance and ensure it is updated with the latest information on security threats and defenses available.

An extensive study of security policies can be found in [FER09].

In this chapter and the previous we have introduced a variety of patterns applicable to Next Generation Networks. We have also looked at specific expressions of these patterns in the form of algorithms, protocols, and tools available, which can be readily used “off-the-shelf” to compose an overall security solution. In Chapter 10 we proceed to define a specific “real life” network, one of the many which are being designed and developed by operators across the globe with IMS as its core, and we design a security solution around it.

PART C

SECURITY CASE STUDY

10 DESIGN CASE – 3G WIRELESS OPERATOR

10.1 Introduction

Wireless operator Alfa Multimedia Telecommunications (AMT) operates a GSM network in North America with 4.5 million subscribers. It has national reach via its own radio access network (RAN) and via partnering agreements with another GSM network operator. It recently won licenses in the top 20 markets in North America to operate in the 3G UMTS bands. With these licenses AMT will be able to offer high speed wireless packet data services such as video streaming, peer-to-peer gaming, and other bandwidth intensive services such as video calling, "see what I see", etc. It also plans to offer internet access for laptops with integrated UMTS radios or by providing users the required PCMCIA cards. Lastly, AMT will also offer its users voice over IP services over packet data, by means of a software application embedded in the wireless handset. This application will also be able to be used for "push-to-talk" services. It is not expected that most existing wireless subscribers switch over to the new service right away, since this would require new devices, but AMT plans to start a new marketing campaign, extolling the benefits of the new network and the new applications available to users. For example, with VoIP over wireless, subscribers will be able to call users at their PC's or mix voice sessions with video or chat sessions. Users will also be able to choose either push-to-talk or a regular connection. Subscribers will also be able to manage their wireless features such as call waiting, call forwarding, and voice mail, on-line. In summary, AMT's users will soon be able to enjoy the benefits of the upcoming Rich Communications Services (RCS) offerings [NOK08].

In order to achieve all this, AMT plans to introduce IMS as the session layer control architecture in the packet core network. Voice over IP wireless calls, and other packet and multi-media applications will be handled by the IMS core. For legacy circuit switched voice traffic, the existing 2G radio access network will be capped at its existing capacity. New subscribers will be provided 3G phones. In order to connect these new subscribers, 3G radio access as described below will be deployed next to the 2G equipment. In the circuit core, the existing network elements will be decommissioned and a new 3GPP Release 4 architecture will be deployed. This circuit core equipment can handle both 2G and 3G access networks. Some manufacturers provide upgrades to their 2G core network elements to be able to handle the 3G radio access network, but this usually just provides a few more years of life to outdated equipment. Introducing a Release 4 core architecture provides many benefits. This architecture is explained in section 10.3.

In time, as 3G phones come down in price, more and more subscribers will be migrated to voice over IP, allowing AMT to decrease expenditures in traditional circuit switched network equipment, and to concentrate more of its assets and new development on packet switched services, which by their nature, can increase ARPU (Average Revenue Per User).

In this chapter we develop the security strategy for AMT. First we start by showing what its current architecture looks like. We then describe in detail the target architecture, including the radio access, the 3G core network, and the IMS core, including back office systems. Next, we will describe the six applications that AMT plans to launch initially on the new network. Finally, we build a security infrastructure around AMT's network by looking at the task in a down-up progression, first by looking at the network elements, then at the user scenarios, followed by a whole-network view, and ending with an overarching security policy.

10.2 Current Network

AMT's existing network is a 2G GSM network, composed of the following logical network elements (the physical distribution of these logical elements is not important and varies from manufacturer to manufacturer:

Core

- Mobile Switching Centers (MSC)
- Home Location Register (HLR)
- Authentication Center (AuC)
- Equipment Identity Register (EIR)
- Visitor Location Register (VLR)
- Gateway Mobile Switching Center (GMSC)
- Short Message Service Center (SMSC)

Radio Access

- Base Transceiver Stations (BTS)
- Base Station Controllers (BSC)

Figure 10.1 shows a diagram of this architecture.

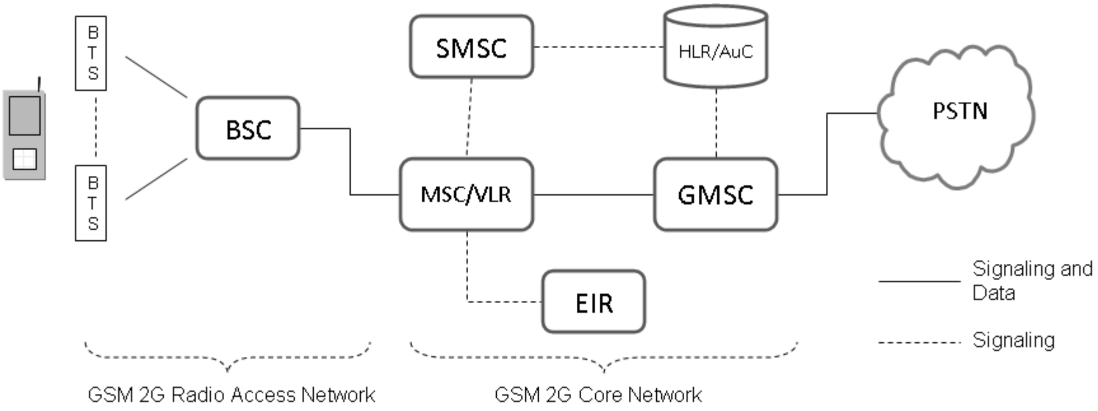


Figure 10-1 - AMT’s Existing 2G Network

AMT’s 2G mobile network does not provide the capability for offering packet switched data services. AMT plans therefore to cap investment in 2G and overlay on top a 3G network as described in the next section. As we shall see, migrating from 2G to 3G does not mean complete new network elements and architecture.

10.3 Planned Network

AMT plans to deploy a UMTS 3G network next to the capped 2G GSM network. In addition, in order to better use the high-speed packet data capabilities of the UMTS radio access and terminals, AMT will also deploy an IMS core network. This IMS network will permit the introduction of the new multimedia applications listed in this chapter.

Upgrading a 2G GSM network entails mainly two activities: deploying the necessary 3G radio access elements usually next to the existing 2G equipment (i.e. in the same radio towers, same cabinets), and upgrading the core equipment (some new modules, new software) to be able to interface to the new 3G radio equipment.

3G Radio Access

The 3G successor to the 2G radio access network is called UTRAN, which stands for UMTS Terrestrial Radio Access Network. The UTRAN enjoys advantages over its predecessor GSM access network in several important categories [BOM02], the main one being that UMTS is a 3G technology, meaning it is designed also for the transport of packet data, and not just circuit data as was GSM. And as would be expected, it also brings with it a considerable increase in the data rate, which can reach up to 2 Mbps, compared with the usual 13 Kbps of GSM. Finally, UMTS also brings improvements in security, in the form of published cipher algorithms (as opposed to unpublished in GSM), a longer cipher key of 128 bits (versus 64 in GSM), mutual authentication to prevent base station masquerading attacks (GSM was vulnerable against this type of attack), and cryptographic protection not just from the handset to the base station as in GSM, but further into the network, to protect the user also from attacks that could occur within part of the operator’s network.

The main elements of a 3G radio access are the base station, called the Node B, and the Radio Network Controller (RNC). An extensive treatment of UMTS is given in [SMI02].

3G Core Network

The architecture of the core in a 3G UMTS network does not vary greatly from that of a 2G GSM network. Other than the new elements needed for the new packet-based services, the network elements for voice services can just get by with upgrading the software to be able to interface to the new radio access components, as well as to be able to control two types of accesses simultaneously (2G and 3G). The amount of necessary upgrades will depend on the individual vendor implementations.

The problem with simply upgrading the 2G core network elements with new software is that as the network grows and more subscriber capacity is added, this legacy core equipment would have to be expanded as well. AMT has chosen not to do this for several reasons: the 2G core network elements are based on outdated technology, they are built on manufacturer proprietary hardware, and they are expensive to operate, maintain, and enhance with new features. In addition, the architecture does not follow the next generation network philosophy of separation of functions. There is however an alternative to upgrading the legacy 2G core network. It consists of replacing it with a 3GPP Release 4 core architecture, which, as seen in the figure below, is IP-based, it separates the control plane from the media plane with open interfaces in between, and is often based on commercial IT platforms, either blade server technology or stacked servers. The figure below also shows the new 3G packet core network elements SGSN and GGSN.

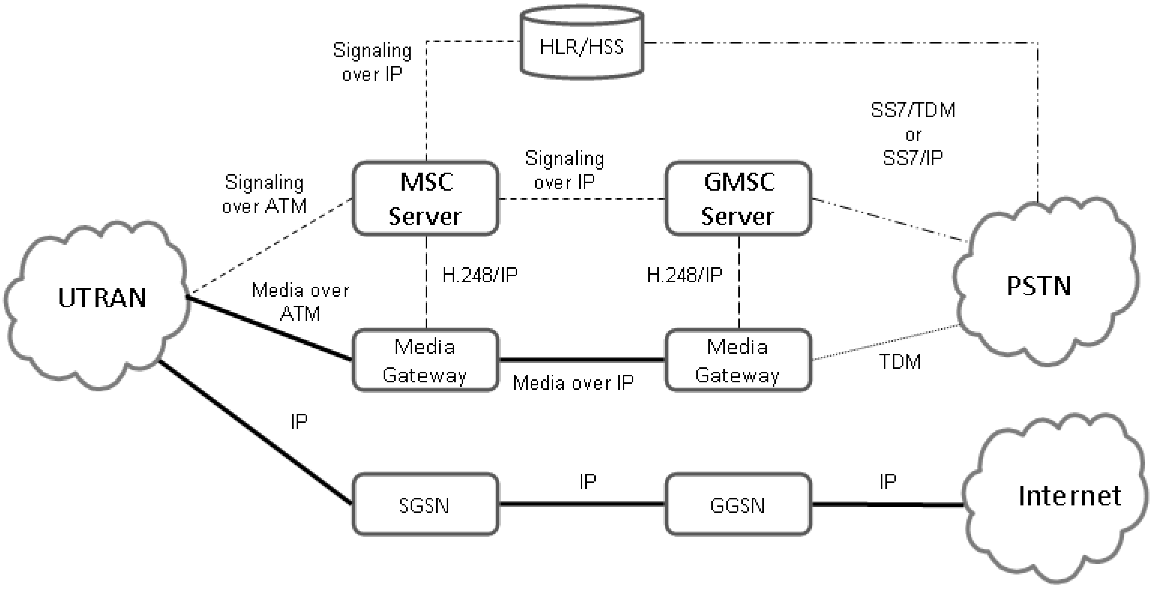


Figure 10-2 - 3GPP Release 4 Core Architecture

IMS core

AMT's initial deployment of IMS will be as an overlay to the existing control architecture. The reason for this is that IMS cannot yet service the existing population of 2G subscribers. It will only be able to do this with the introduction of the Mobile Access Gateway Function (MAGF) in a future 3GPP release (planned for Release 8). Until then, the legacy 2G users will be serviced by the core elements shown in the previous figure.

The IMS core will be used for the added services listed in section 10.4. AMT will offer these premium services to those subscribers who want them and who want to upgrade to 3G handsets or PDA's. These subscribers will access the IMS, and these new premium services over the packet core elements in Figure 10.2, the SGSN and GGSN. The IMS core will comprise the following elements:

- Home Subscriber Server (HSS)
- Call Session Control Server (CSCF)
- Breakout Gateway Control Function (BGCF)
- Media Gateway Control Function (MGCF)

- Policy and Charging Rules Function (PCRF)
- Core-Border Gateway Function (C-BGF)
- Interconnect-Border Gateway Function (I-BGF)
- Voice Application Server (VAS)
- Media Resource Function (MRF)
- Push-To-Talk Server (PTT)
- Location and Presence Server (L&P)
- ENUM/DNS Server

In addition, AMT has decided to start consolidating its database services of all types onto a single geographically distributed database store (DBS). The DBS is used for raw storage of subscriber and other data, and it can serve multiple front-end applications via a common interface, typically LDAP. This greatly simplifies operations by unifying the administration and provisioning of all kinds of data onto a single platform. This platform can then satisfy the data storage and retrieval of information from multiple sources via standardized interfaces. Furthermore, the DBS is based on in-memory storage technology, with periodic back-up to disk. This is needed in order to satisfy the stringent millisecond time access requirements of telecommunications applications. In the initial deployment, the DBS will serve the front end applications: HSS, HLR and PCRF. In the future, it is planned to also decouple the VAS from its subscriber database by offloading it to the DBS.

Not listed but also part of AMT’s basic network is an MPLS backbone connecting all its core sites and radio access points of presence.

Figure 10.3 shows AMT’s overall planned network architecture.

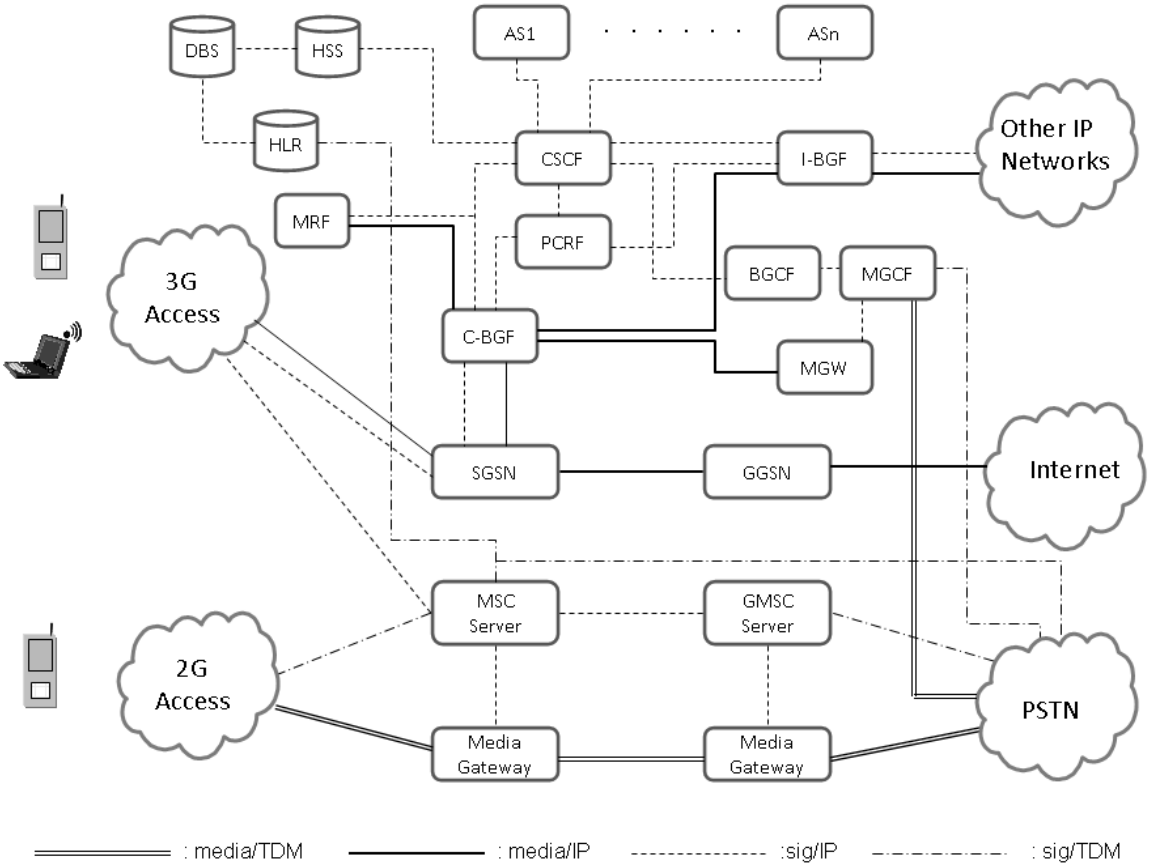


Figure 10-3 - AMT’s Overall Architecture

In the figure above we have combined the three main elements of AMT’s core network into a single diagram: the upgraded 2G GSM core at the bottom, the 3G Packet core in the middle, and the IMS core, applications and database store and front-ends at the top. The interfaces between each network element are marked as carrying either media or signaling, and whether at the link and network layer they use TDM or IP. This will be important later since in this work we are only concerned about NGN security (i.e. the protection of the IP elements and interfaces).

Back Office and Auxiliary Servers

For clarity, we have left some network elements out of this diagram: operations back office elements like element managers and billing servers, the backbone IP routers, DNS and ENUM servers, and firewalls. We show these in Figure 10.4 below (minus the IP routers since these are not relevant to our discussion):

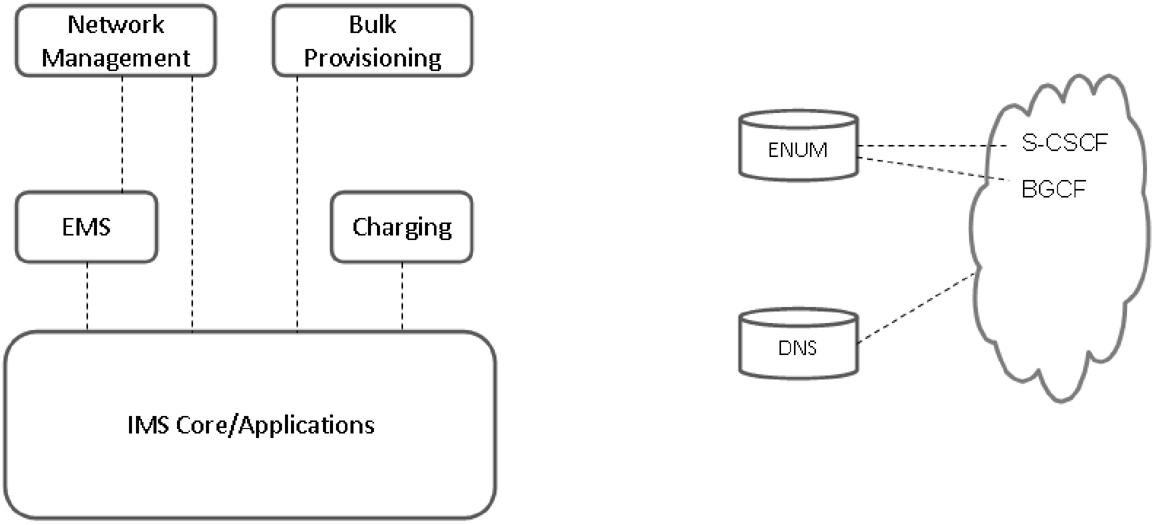


Figure 10-4 - AMT's Back Office and DNS/ENUM Servers

The EMS (Element Management System) is used for the day-to-day administration, operations, fault reporting and management of every IMS core element, and possibly some of the application servers. The Charging server is used for offline charging (AMT uses only offline charging, no online or pre-paid charging). Relevant charging records are sent to this server by core elements and application servers either as charging events occur, or periodically, depending on AMT's preferences. The Network Management server is used by AMT to gather, monitor, and use performance data on individual elements to coordinate overall network resources. The Bulk and Flow Through Provisioning element is used to effect subscriber data provisioning, either in bulk mode, or individually as additions, deletions or changes occur.

All these back office servers need to interface with one or more IMS core network and applications servers. We will address the security needs of these elements later when we put in place a security solution for this network.

The DNS and ENUM servers are inside the core domain itself, and need to be able to be addressed by certain network elements within the core. Because of AMT's architectural decisions regarding routing, we know that the ENUM server will only be used by the S-CSCF and the BGCF. Similarly, the DNS can be queried also from the core, but from a larger number of network elements. These rules will have to be part of the overall security solution engineered in Chapter 10.

CSCF Decomposition

Also for clarity, the different types of CSCF (I, P, S) have been shown in a single box. For reasons of scalability and reliability, AMT will deploy them as separate hardware elements. In addition, We will also assume that the function which controls the x-BGF, the Border Control Function (BCF) is capable of residing in the P-CSCF and I-CSCF. This is the case with some IMS vendors. Finally, the BGCF is being shown as a stand-alone element although most vendors, including the one selected by AMT, can also include it as part of the CSCF functionality. If this additional detail is pictured and the non-IMS parts are removed, the network looks as shown in figure 10.5

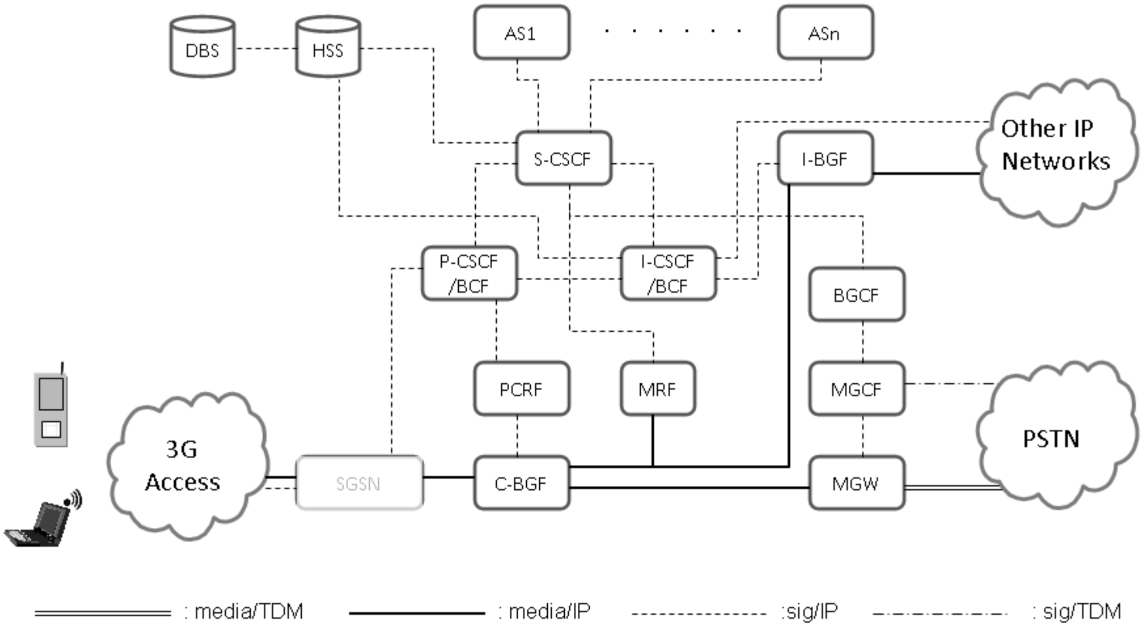


Figure 10-5 - AMT's IMS-Only Architecture

10.4 End User Applications

We now look at what new applications are being deployed over this new control core. AMT plans to attract subscribers from its plain wireless telephony service to the new and more revenue generating IMS-based service by introducing brand new applications over wireless packet access. There are no limits to how many applications can be developed by mixing: voice, text, video, location information, presence information, user profiling, etc. AMT has decided to start with the following set of 6 applications:

1. Voice over IP calling (handset and laptop pc)
2. Amateur Reporter (1-way video with 2-way speech)
3. Push-to-Talk
4. Video telephony
5. Location based Personalized Information Pull
6. Location based Gaming

We give below a summary description of each use case, including diagrams showing the path that the application signaling takes through the core network. After describing all the end-user scenarios, we will design the security architecture.

10.4.1 Voice over IP Calling via packet core

A voice over IP call from a packet data capable (2.5G or 3G) wireless device is at the application level essentially no different from a VoIP call made via the home cable or DSL connection. A SIP User Agent application must be installed on the mobile device. After the client has obtained an IP address from the network provider, it registers with the IMS registrar service (the S-CSCF) using a pre-programmed FQDN. SIP is used for all signaling between the user and the network. The SIP REGISTER message contains the user's private ID, public ID, and source IP address. After a valid registration, the user is ready to accept or initiate voice calls.

Although for the end-user there are no noticeable benefits from the call going over IP instead of TDM, the carrier does benefit from the flexibility and lower cost of IP facilities. The end-user will benefit when, due to the flexibility of IP, the voice call can be combined with other IP services to build richer applications.

Figure 10.6 below shows the path that the call signaling messages would take, for a call either to another mobile device on the same network, to another IP network, or to the PSTN.

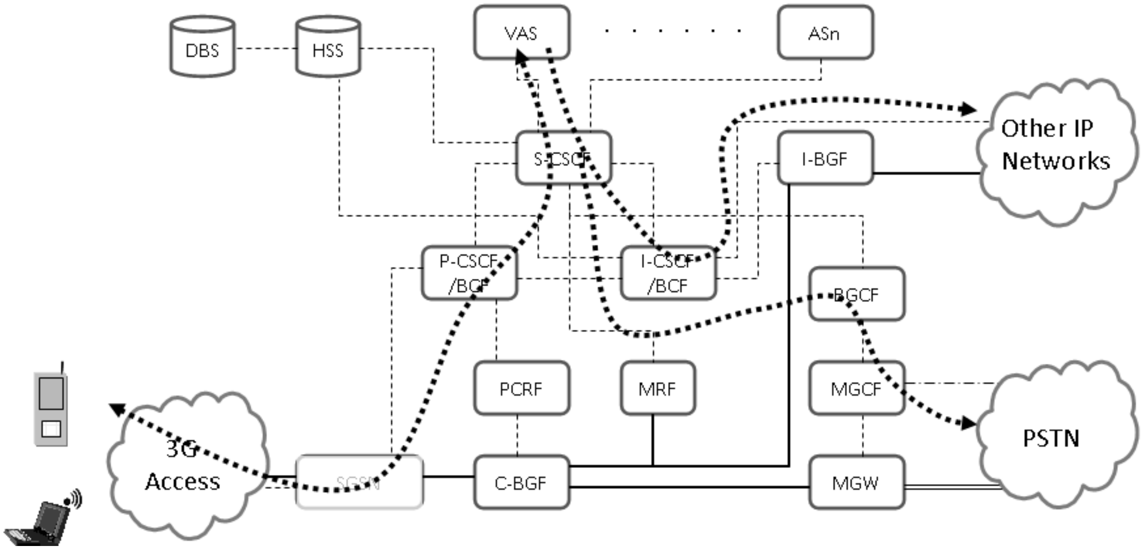


Figure 10-6 - VoIP Call Flow

The call originates at the Mobile Terminal and traverses the UMTS Radio Access Network (UTRAN) to the first IP element, the SGSN. The first SIP aware network element is the P-CSCF, which passes the SIP messages on to the S-CSCF, and this one in turn to the Application Server (AS), in charge of the call logic. The signaling returns the same way if the called party is also a mobile terminal or it travels via the I-CSCF (for topology hiding, this will be explained later) if the call terminates in another IP network, or it traverses the BGCF and MGCF if going to the PSTN or PLMN. There will be intermediate side queries or control messages (not shown) to the ENUM server, the HSS, the PCRF, the MGW, and both BGF's at different points in the call. There may also be billing messages to the billing server. All these internal communication paths have to be considered for the security design.

The media path is much simpler and is indicated by the solid lines. If the call is to another wireless subscriber, the media will simply get looped at the C-BGF. If the call is to another IP network, it will traverse from C-BGF to I-BGF. And if to the PSTN/PLMN, it will go out via the MGW. In case of set up problems where an announcement has to be inserted, the media is directed to the MRF.

10.4.2 Amateur Reporter (1-way video with 2-way speech)

This feature, sometimes also called “see-what-I-see”, involves a wireless user with a camera phone and a broadband data connection. In a typical scenario, the cellular subscriber establishes first a voice connection with another subscriber, either wireless or fixed, but also with data services and a video client on the device. Depending on the design of the client on the mobile terminal, the voice connection may be via the circuit switched or via the packet switched domain, i.e. via traditional GSM or via IMS. In AMT’s case, all connections will be IMS-based. After the voice path is established, the originator of the call initiates a one-way video session from the camera phone. When complete, two simultaneous media paths exist from the sender to the receiver, a bidirectional one for the voice way and another for the one-way video.

Figure 10.7 shows the path through the network for the signaling, including the necessary setting of policy for the speech and video flows by the PCRF on the C-BGF.

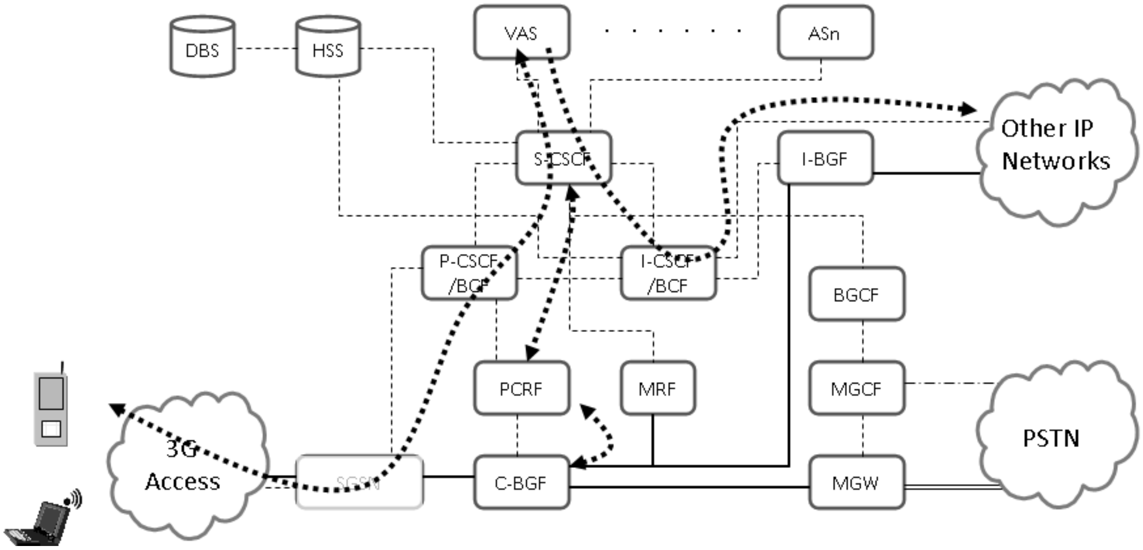


Figure 10-7 - “Amateur Reporter” Signaling Flow

10.4.3 Push-to-Talk over Cellular (PoC)

This feature aims to emulate the popular “walkie-talkie” type of communication commonly used for limited distances via point-to-point or point-to-multipoint radio, except that this service will be for coast-to-coast calls. There is an existing circuit switched technology in use to provide this service, marketed by Sprint. It uses IDEN (Integrated Dispatch Enhanced Network) [SMI02]. In IMS, unlike in circuit switching, there is no dedicated channel per communication session, which can have quality of service ramifications if the network is not well designed. On the other hand, IMS adds other capabilities when service enablers like location, presence, etc. are used to enhance the application. Unlike the previous features, Push-to-Talk will usually be designed and deployed in a separate application server, the Push to Talk over Cellular (PoC) server which adds complexity to the message flow. We also see the Location and Presence (L&P) server coming into play. This server may be one or two separate physical applications, depending on the vendor. It is responsible for maintaining information about subscriber location and presence (i.e. ability and readiness to engage in a PoC session). It also maintains the group or “buddy” list associated with the PoC service. An originator of a PoC session can select from his or her buddy list one or more buddies with whom to participate in a session. This means the session can be one to one, or one to many. The L&P server provides the PoC server with the presence information relevant to all the chosen participants.. The 3GPP defines the architecture to support OMA (Open Mobile Alliance) Push-to-Talk in TS 23-979 [3GP979]. The OMA PoC architecture is defined in [OMA01]

Figure 10.8 shows the signaling path through the network.

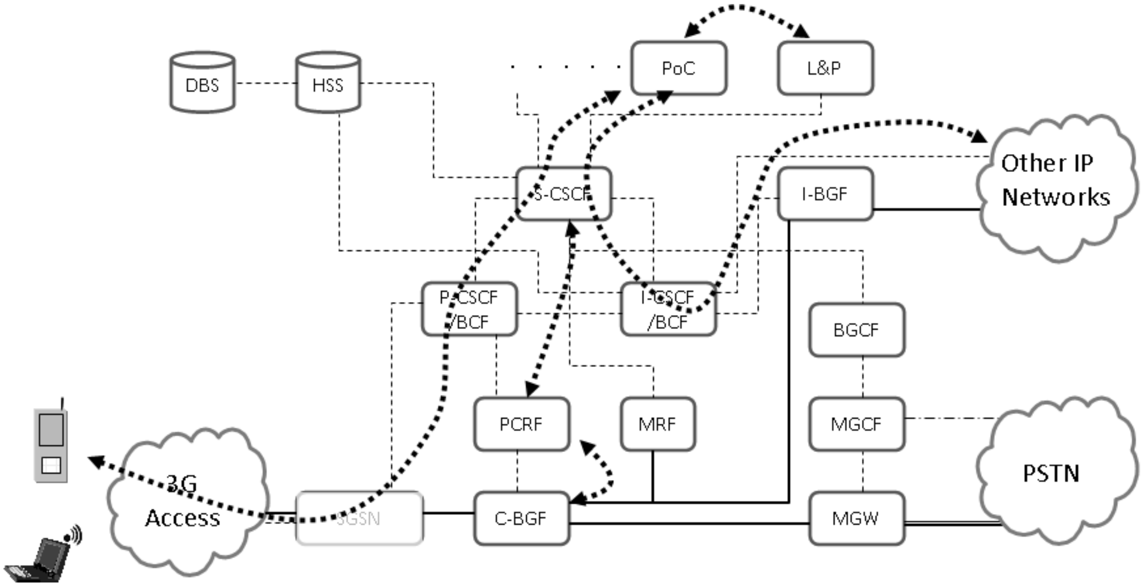


Figure 10-8 - “Push to Talk over Cellular” Signaling Flow

This scenario has two main differences with respect to VoIP, the first application we saw. Firstly, there is a new function, the L&P, separate from the application server itself (the PoC server), and which is mainly a data repository with some added control functions like the capability to accept subscriptions to data changes, and to send data change notification messages. It is used by the main application (PoC) as data source for finding the status and settings of the subscribers using the Push-to-talk service. In that sense, it may have different security requirements from that of a control server (see chapter 5). Secondly, the media flow in Push-to-talk is very different from that for VoIP, in that it is discrete, half-duplex, uses the MSRP protocol [RFC4975] for transferring media bursts, and traverses the PoC Application Server (see Figure 10.9). PoC does not interwork with the circuit-switched PLMN/PSTN.

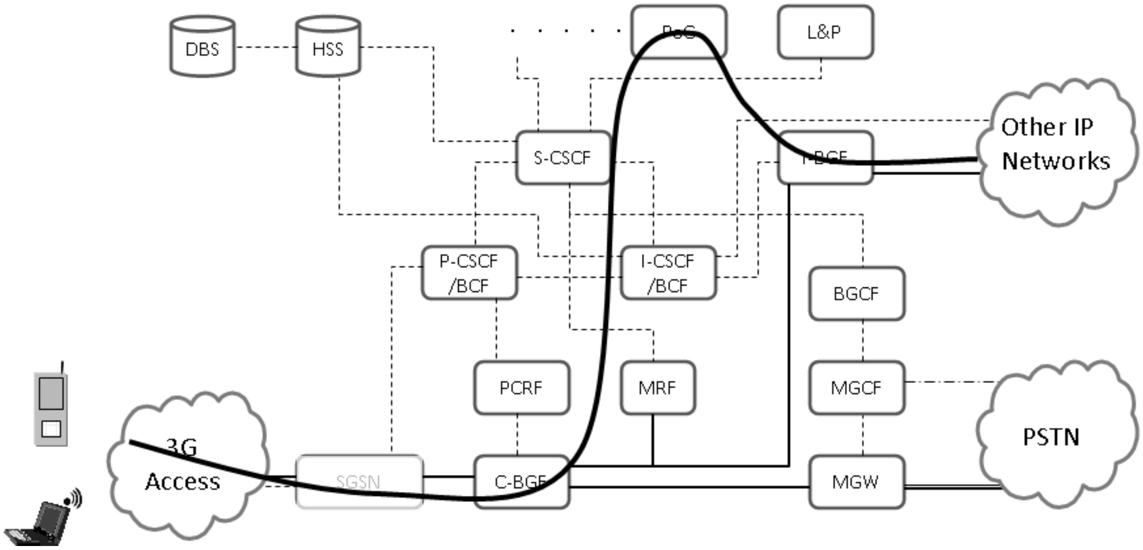


Figure 10-9 - Media Flow for PoC application

10.4.4 Video telephony via packet core

With video telephony over packet core, two cellular users with 3G mobile terminals equipped with video camera can hold a video session. Both packet flows, for video and for speech, will traverse the network via the IP backbone, as opposed to other possible ways of deploying this service, in which for example the speech connection could be set up via the circuit switched network using traditional GSM procedures.

It is possible to also provide this service such that the 3G cellular user can hold a video session with a fixed IMS subscriber which registers from a lap top computer, equipped with a camera and an IMS SIP client. With both cases, the video and speech streams can be exchanged between both users without the need of going through a trans-coder (a function which translates between to different ways of encoding analog signals), as long as both UA's can negotiate a common codec. There are other possible types of originating-side/terminating-side combinations where trans-coding would be necessary, for example, if in the use cases above the UA's do not share a common codec, or where one of the end points is a traditional circuit switched connection (PLMN or PSTN), in which case, an IP/TDM video gateway is required.

With the launch of this service, AMT will only support video sessions between own IMS subscribers attached to the own radio access network or roaming in a partner's network. In that case, the network elements involved and the message flow are the same as those for a regular VoIP call.

From an architectural, signaling, and media traversal point of view, there are no differences between this application and plain VoIP. The network elements taking part in both services are the same. The only difference is the existence of a second set of media flows for the video.

10.4.5 Location based Personalized Information Service

Like the previous application, this service is based on two primary "enablers": presence and location, provided in AMT's case by the L&P server. The use case for this application could be the following: a mobile 3G user who travels frequently can call up an application on his or her mobile device and register his or her interests, for example: Indian restaurants (only 4 stars or better), museums (modern art), jazz clubs. Naturally this can also be done from a PC via a web application. The application server ("Info" in Figure 10.10) stores this information. When the user travels outside of his or her local area, and the handset is turned on, the SIP user agent registration in a new city is communicated to the L&P server by the registrar (the CSCF) via an IMS mechanism called 3rd party registration, and the L&P in turn sends this information to the application server. This registration message contains the location of the user, either based on information supplied by the mobile terminal (i.e. GPS), or on information supplied by the radio access network (i.e. cell ID). The application can now search a database or the Web for entries related to the interests of the user, and can "push" them to the mobile terminal application in the form of an HTTP link. The designers of the user application can design different options into the application interface such as: show the information immediately, store it for later retrieval without alerting the user, poll the server regularly for updated content, push information only when in selected ZIP codes, allow click-to-dial from the HTTP page, etc. Similar types of applications are already in use by the internet search engines, in combination with map and direction applications.

This application brings in an HTTP server into the configuration, located on the Info server. The handset or PDA client must also have HTTP capabilities. Having an HTTP client in the operator's network, which can interface with some of the IMS network elements, brings in different security risks which will have to be safeguarded against.

The following figure shows the path through the network for the application.

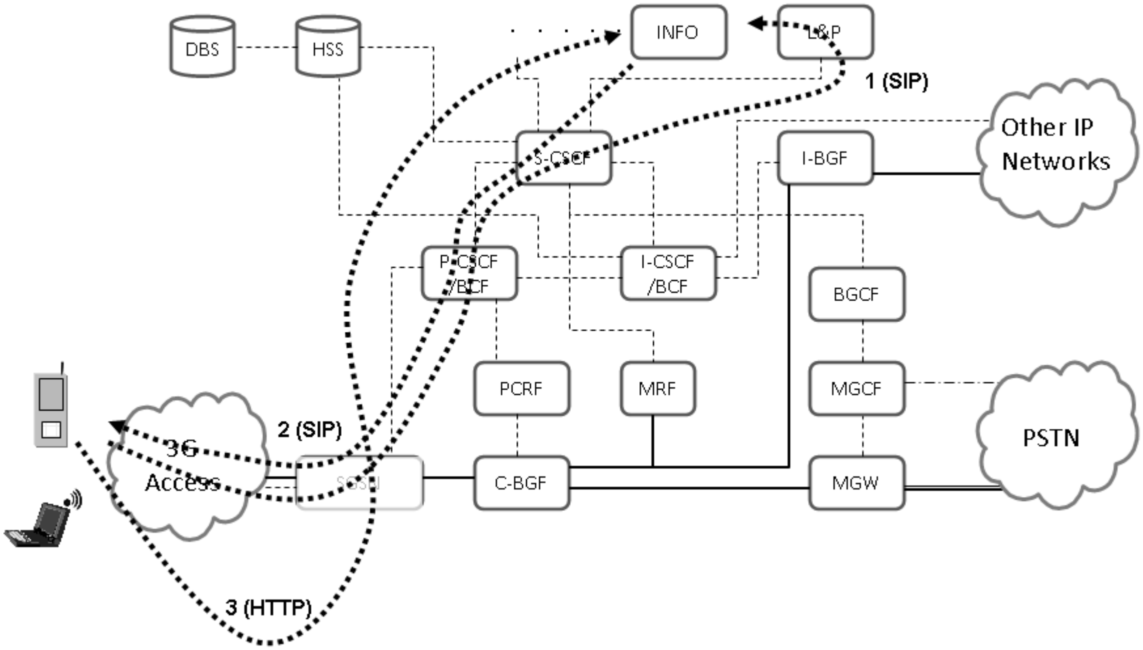


Figure 10-10 - Location-based Personalized Info

We can see that the first two dialogs in this application are SIP-based, labeled 1 and 2. the first one is when the subscriber registers in a new location and this information is shared by the IMS core with the Location and Presence server, and this one, with the Info application server. The second one is when the Info server pushes a link to the user which points to the information based on the preferences found in its database. The application in the handset can then request this information by means of HTTP. The Info server must then have both SIP and HTTP capabilities.

10.4.6 Location based Gaming combined with Group Lists

As with any application which involves divulging one’s location and preferences, the application just described, and Location based Gaming described here, can be abused by internet predators or even physical ones. Safeguards must be designed into the application (and not just the minimum security features that will be described later), and common sense must be exercised by the user, in order to avoid potentially dangerous situations.

AMT’s location based gaming will provide the enabling components for developers to write, prototype, and launch multi-player games. The possibilities are limitless: imagine a game, it can be something as trivial as chess or checkers, or something more complicated such as role-playing games, or scavenger hunts. These games, which are played every day by millions of people connected to the internet, usually lack real human interaction, sometimes for good reason. They are usually based on group lists, also called buddy list or communities of interest. A buddy list allows a user to register with an internet based application, and via icons, “publish” user relevant information such as presence (“in” or “out”), mood (“in, but leave me alone”), relevant interests (“I only want to play this version of the game”), device capabilities (text, speech, video). Players can then invite each other to take part in game sessions. Location based gaming adds the variable of location. For example, you land in New York and when your mobile device registers, the application notifies you that there is another chess player of similar rating in the city, and that he is willing to play a match. You engage him in a match and at the end decide to meet at a local Starbucks to review it play by play. This is a simple use case. More complicated ones could involve some sort of “tag” or “treasure hunt” where the players would be guided through the city to “collect” clues or items, followed by a group gathering at some venue in the city.

AMT will start by introducing three such games, and then open up the network interfaces to allow third-party providers to introduce their own applications to AMT’s users.

The enabling components needed for such games will be the same as for the previous application and as shown in Figure 10.10 Obviously, the gaming server itself (in place of the Info server), which will have access to a presence and location server with the capability to store and manage buddy lists, and depending on the game, there may be a need for an HTTP server as well.

10.5 Security Design

Now that we know the architecture and the use cases we can start designing the security components. The preceding use cases use every network element in AMT's IMS core, IP core, access network, and interconnect network. Putting in place a security solution that covers these 6 cases, will very likely cover any other scenarios that can be developed. Using the methodology discussed in Part B, where we abstracted the NGN and ascribed one of four different roles to every possible network component, and the security tools and methods described in Chapter 9 (as particular examples of the patterns discussed in Chapter 8), we design a security solution for AMT which defends against the possible threats listed in Chapter 7.

Specifically, we will arrive at the security design by taking the following steps:

1. Classify the network elements into one of the four types introduced in chapter 5.
2. Consider the relevant threat or vulnerability and apply the matched security measure
3. Examine any possible characteristics the abstracted model may have left out with respect to the network element in question

We will then look at the designed solution from the point of view of the 6 scenarios being considered and the overall solution. This gives us the final 2 steps:

4. User scenario specifics
5. Whole network view

10.5.1 AMT's IMS Core NE Classification

In Figures 10.11 and 10.12 below we show again for convenience AMT's IMS Core network with all its components.

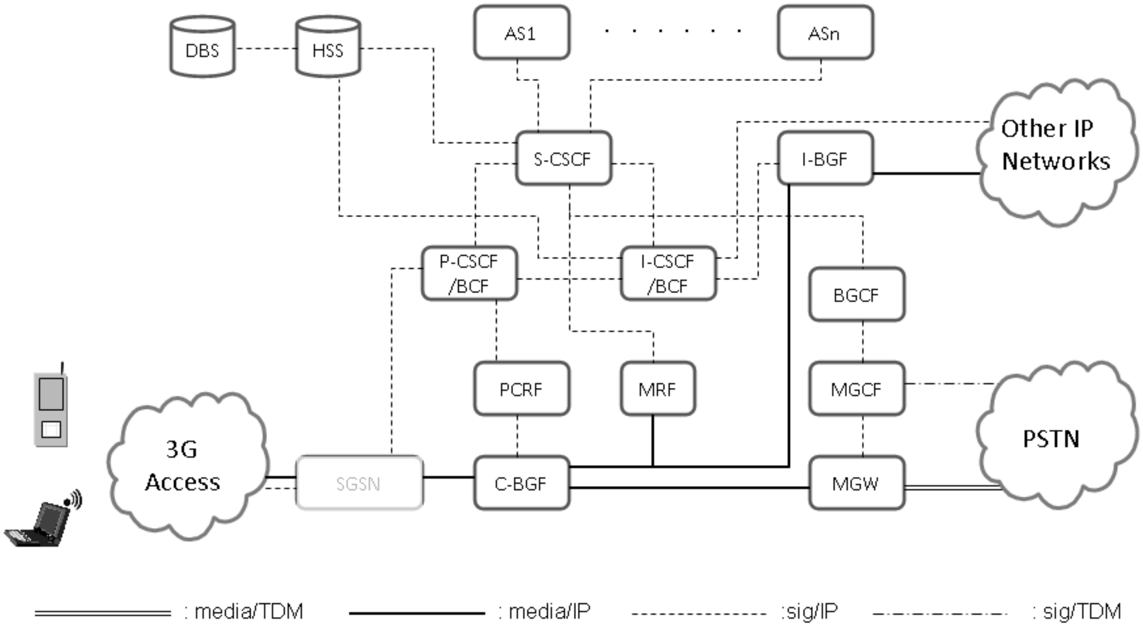


Figure 10-11 - AMT's IMS Core

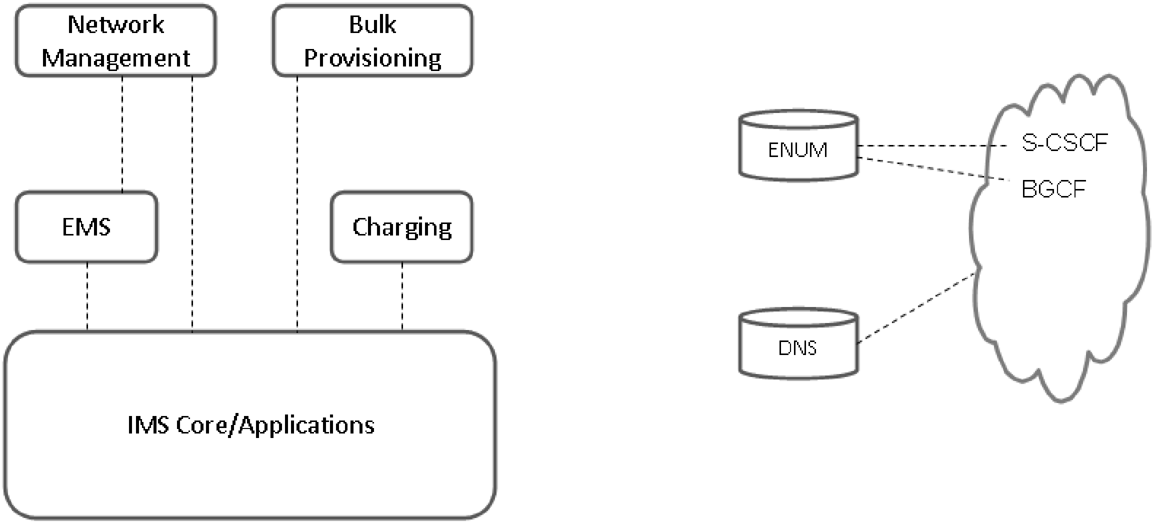


Figure 10-12 - AMT's IMS Core Back Office and DNS/ENUM servers

We recall from Chapter 5 that the four possible types of network elements are: Database servers, Control/application servers, Media routers/gateways, and Sentinel Gateways. From the diagrams above and the individual use-case scenario diagrams we list all the network elements in AMT's IMS core:

P-CSCF/BCF (Proxy Call/Session Control Function/Border Control Function)

I-CSCF/BCF (Interrogating Call/Session Control Function/Border Control Function)

S-CSCF (Serving Call/Session Control Function)

HSS (Home Subscriber Server)

DBS (Database Server)

PCRF (Policy and Charging Rules Function)

MRF (Media Resource Function)

I-BGF (Interworking Border Gateway Function)

A-BGF (Access Border Gateway Function)

BGCF (Breakout Gateway Control Function)

MGCF (Media Gateway Control Function)

MGW (Media Gateway)

PoC AS (Push to Talk over Cellular Application Server)

VAS (Voice Application Server)

INFO Application Server

Gaming Application Server

L&P (Location and Presence enabler)

ENUM (Telephone Number Mapping server)

DNS (Domain Name System)

EMS (Element Management System)

Charging server

Network Management server

Bulk Provisioning server.

We classify AMT's network elements along these types as follows:

Database servers: DNS, ENUM, DBS

Control/Application servers: P-CSCF/BCF, S-CSCF, I-CSCF/BCF, HSS, PCRF, MRF, MGCF, PoC AS, VAS, INFO AS, Gaming AS, L&P, EMS, Charging server, Network Management Server, Bulk Provisioning server

Media Routers/Gateways: I-BGF, A-BGF, MGW,

Sentinel Gateways: P-CSCF/BCF, I-CSCF/BCF

Notice that there are two functions, the access network Border Control Function and the interworking Border Control Function, which are categorized both under Control/Application and under Sentinel. The reason for this is that the vendor chosen by AMT has integrated these sentinel type functions into the P-CSCF and I-CSCF respectively.

10.5.2 Vulnerabilities and Defense Measures

We now use the conclusions from chapters 5, 8, and 9 to list the important information that will help us in selecting the right tools and methods to apply to each of AMT's network elements. First, we summarize this information in a table for each NE type, and then we decide what specific measure related to the given pattern to propose as part of the of the security design. This is shown in Table 10.5.2.1 below.

10.5.2.1 Dababase NE Types

NE Type (Chapter 5)	Relevant NE's	Vulnerability (Chapter 7)	Security Patterns (Chapters 8&9)
Database Server	DNS, ENUM, DBS	- Information theft - Data corruption - Denial of service	- Data Confidentiality - NE Comm. White List - Strict Protocol Adherence - NE Hardening

Table 10-1: Database Server Vulnerability and Patterns

From the table we can now proceed to define specific tools and measures which derive from the fourth column.

Specific IMS Security Tools

Data Confidentiality: encryption must be used for data in any of these network elements which can be displayed via the Element Management System (some data cannot be displayed but can only be requested from another non-EMS element), to prevent unauthorized use. Typically this will only affect the DBS and the subscriber data it contains. DNS and ENUM servers do not contain such data.

NE Communication White List: all database NE's must be allowed to communicate only with a given set of IMS core elements. This white list must be compiled in advance for each NE by the operator and the network integrator. The white list can be a set of IP addresses, or for more flexibility, a set of FQDN's. A white list will prevent unauthorized access to the database servers from servers other than where the information is needed. The operator must ascertain during the product requirements definitions phase that the NE vendors have this capability in their products.

Strict Protocol Adherence: the vendor must specify during the acquisition phase which protocols and which extensions must be satisfied. A key protocol for these servers is Diameter. Diameter is extensible, in that different applications require implementation of different AVP's (Attribute Value Pairs). Only those AVP's which are relevant to the application should be expected by the database servers. Others should be rejected gracefully or ignored. Other protocols used must be equally specified. This requirement should apply only to those servers which can be queried from elements outside the control of AMT, in this case, DNS and ENUM, since those elements in AMT's domain will not attempt protocol attacks.

NE Hardening: the database servers in question are intense use applications. Almost every user action will cause access to one or more of them. Some of them, like DNS, are also very attractive targets for DoS attacks. The operator must require hardening of these three NE's during the product specification phase. It must request proof of the steps taken, including: operating system processes disabled, sockets closed, protocols disallowed, hardware modules removed or not populated, other physical tampering measures taken, security software pack level, and root passwords. It must also request a description of the vendor's plans for continuously monitoring the discovery of new security flaws, and a guarantee that steps will be taken to close them within an adequate length of time.

10.5.2.2 Control/Application Server NE Types

NE Type (Chapter 5)	Relevant NE's	Vulnerability (Chapter 7)	Security Patterns (Chapters 8&9)
Control/Application Server	P-CSCF/BCF, S-CSCF, I-CSCF/BCF, HSS, PCRF, MGCF, PoC AS, VAS, INFO AS, Gaming AS, L&P, EMS, Charging server, Network Management Server, Bulk Provisioning server	- Information Theft - Denial of service - Theft of services - Service destruction - Data corruption - Vehicle to attack other networks	- Data Confidentiality - NE Comm. White List - Strict Protocol Adherence - NE Hardening - Topology Hiding - Time Limited Registration - Authentication - Authorization - Communication Confidentiality - Communication Integrity - Non-repudiation

Table 10-2: Control/Application Server Vulnerability and Patterns

From the table we can now proceed to define specific tools and measures which derive from the fourth column.

Specific IMS Security Tools

Data Confidentiality: the information theft vulnerability is a serious threat only with respect to the NE's within this group which hold subscriber or operator information (and have considerable logic to process it, that's why they are in this group). Out of the list in the table, those would be the L&P, EMS, Network Management, and Charging servers. Out of these four, three are in the "back office", i.e. not really part of real time service delivery, and one of them is in the applications layer. The type of data they hold is subscriber "permanent" data, transient data, location data, service related data (e.g. preferences), charging data, some topology data, and probably some operator data. Confidential data which cannot be read out or displayed via the EMS, will have no need of encryption, as long as, when it is transmitted from element to element, communication confidentiality is guaranteed. If confidential data however could be displayed at an EMS, encryption should be provided. The operator and integrator should jointly decide what data, transient and permanent, needs to be encrypted.

Network Element Communication White List: like with the previous NE type, not every element in this list has a need to communicate with every other element. Likewise, access from other elements not in the list must also be restricted. As an example of what needs to be done here, we show below the white list for the S-CSCF, PCRF, and MRF:

Network Element	Allowed Communication Partners	FQDN
S-CSCF	*P-CSCF* *I-CSCF* *HSS* *VAS* *L&P* *INFO* *MRF*	*pcscf@AMT.net* *icscf@AMT.net* *hss@AMT.net* *vas@AMT.net* *locandpres@AMT.net* *infoas@AMT.net* *mrf@AMT.net*
PCRF	*P-CSCF* *A-BGF*	*pcscf@AMT.net* *abgf@AMT.net*
MRF	*S-CSCF*	*scscf@AMT.net*

Table 10-3: White List table for S-CSCF, PCRF, MRF

Strict Protocol Adherence: the most common protocol within this category of NE's is SIP, followed by Diameter. Others used are SNMP, MGCP (or H.248), and HTTP. Like for the previous category, the operator needs to request during the product specification phase that this measure be adhered to. In particular, widely known protocols such as HTTP, must be rigorously implemented. If necessary, the operator can request proof that the protocol stack has been tested against various protocol layer attack tools. This requirement only applies to elements that communicate with external domains. One very useful feature that can be requested by the operator with regards to any SIP network elements is that they contain a configurable table of what Methods shall be accepted by the server. In this way, flexibility is achieved for this security measure, since SIP Methods can be removed from the list as they are needed due to new applications.

Network Element Hardening: the same requirements apply as for the database server NE type.

Topology Hiding: there are mainly two ways to implement this security measure. The first one is architectural and the second one at the protocol layer. The goal is to hide the details of AMT's internal network configuration (IP addresses, FQDN's, connection matrix) from possible external snooping, either the access side, the interconnect side, or even the application layer. This can be done architecturally in that the operator uses only one single interface towards any of the three domains. Every other NE is placed inside. From the protocol side, this single point of contact must then strip relevant information from all outgoing messages. The operator must request this capability from their border NE vendors at product specification time, by demanding compliance to the relevant sections in 3GPP standards TS 23.228 [3GP228] and 24.229 [3GP229], and then configure the network following these guidelines. The NE's involved are: I-CSCF, P-CSCF, and S-CSCF. This measure will be complemented later with the DMZ pattern, when we look at the network holistically.

Time Limited Registration: this measure pertains in AMT's case to two different entities. The first one is subscribers, which must register with the IMS by means of a SIP REGISTER message to the S-CSCF. The second is network personnel, who must log into the EMS and other servers (Network Management) in order to operate the network. In both cases there needs to be a time limit set, during which the registration is valid. After the set period, a re-registration with authentication must take place. The period lengths must be configurable by the operator.

Authentication: this measure has relevance and must be implemented in the following network elements: S-CSCF – this is the Registrar in an IMS core. Any subscriber wishing services must register with the Registrar first. The 3GPP standards have made several authentication mechanisms available to operators to choose from. In particular, IMS AKA, and Digest AKA Authentication, mentioned in Chapter 9, are both recommended because they provide not only subscriber but mutual authentication. The decision about which method to use will depend on the capabilities of the end-devices available to subscribers. They will also need to support the same method.

Info Application Server – the operator may decide to implement authentication on this server so that only the subscribers who receive an HTTP link can actually retrieve the information, see Figure 10.4.5.1. This can be accomplished by requiring an additional authentication dialog between HTTP server and user, or more efficiently, by implementing the Generic Bootstrapping Architecture referenced in section 9.7. which is a method for providing single-sign on capabilities.

EMS – all means of administration of network elements, back office servers, firewalls, routers, etc. must be password protected.

Authorization: after authentication, authorization grants access to the authorized entity only to those services, features, and privileges for which the subscriber has paid, or necessary for operations personnel to perform their job. In the case of subscribers, the S-CSCF performs this function after downloading the user profile from the HSS. Simple adherence to 3GPP standard 23.228 will satisfy this requirement. In the case of operations personnel, implementation of the Role Based Access Control and Reference Monitor patterns (see section 9.8) will be necessary. Implementation of RBAC and RM in the EMS will have to be a requirement to the vendor during the product definition phase.

Communication Confidentiality: in terms of this defense measure a distinction needs to be made between confidentiality of signaling, and confidentiality in media transmission. Protecting the signaling messages between user and IMS core, among IMS core network elements, between IMS core and application servers, between IMS core and another IP network, and between all network elements and the EMS, is of critical importance to safeguard sensitive user information, topology information which could be used for attacks, as well as critical data such as billing records. On the other hand, providing media confidentiality is not as critical to the operator, although in certain cases it may be desired by end users. Media confidentiality is *not* provided currently in existing fixed telecommunications networks (although it is in the air portion of some mobile networks), and where it's needed, end-to-end cryptography can be used. Providing media confidentiality within the core network would complicate compliance with lawful interception requirements (section 2.4.1), injection of recorded announcements, and connection to voice mail servers, while providing very little added risk avoidance.

To provide signaling confidentiality, we need to look at each of the domains mentioned above:

User to IMS Core – for the use cases to be deployed by AMT, 3G devices will be used. As stated, the 3G UTRAN already provides encryption between the mobile terminal and the access network. No additional measures needed. If in the future, AMT adds user cases which include providing service to fixed subscribers, or to mobile terminals over WiFi, then IMS AKA with optional cryptography, or TLS, will have to be specified.

Intra IMS Core and from IMS Core to EMS – signaling among core elements, including EMS, must be encrypted to protect the network from inside attacks. This will be accomplished by requesting that all elements satisfy reference point Zb as shown in section 9.10, and then configuring the network elements to establish security associations with each other prior to entering live operation.

IMS Core to Applications – this interface should also be protected via the Zb reference point as above.

IMS Core to other IP Networks – signaling between AMT's IMS core network and other IMS or IP networks must be protected. This will mean inserting Security Gateways as per reference point Za in section 9.10. Obviously all of AMT's peering partners will have to comply to reference point Za as well.

Communication Integrity: signaling integrity from the users to the network and vice-versa, in order to thwart man-in-the-middle attacks, is provided by the methods employed also for authentication: IMS AKA or Digest AKA.

Non Repudiation: in the context of the services offered by the operator, non-repudiation between network and user is assured by the mutual authentication methods used.

10.5.2.3 Media Routers/Gateways NE Types

NE Type (Chapter 5)	Relevant NE's	Vulnerability (Chapter 7)	Security Patterns (Chapters 8&9)
Media Routers and Gateways	MGW, MRF	- Denial of service - Vehicle to attack other networks	- NE Comm. White List - Strict Protocol Adherence - NE Hardening

Table 10-4: Media Routers/Gateways NE Types Vulnerability and Patterns

From the table we can now proceed to define specific tools and measures which derive from the fourth column.

Specific IMS Security Tools

Network Element Communication White List: it can be seen from Figure 10.5.1.1 that the two elements under this category, the MGW and the MRF, can only exchange signaling or control messages with other control elements within the IMS core, including but not shown in the figure, the EMS. They will have then only two entries each in their white list: the S-CSCF and the EMS for the MRF, and the MGCF and the EMS for the MGW. For the media a similar white list needs to be created to limit media streams to and from the A-BGF, I-BGF.

Strict Protocol Adherence: only the MGW is exposed to network elements outside AMT's control, since it interfaces with the PSTN/PLMN. The danger of being targeted for protocol level attacks at the internal interfaces is minimal. Even so, during the product specification phase, it needs to be requested that both MGW and MRF be safeguarded against malformed messages. On the PSTN/PLMN interface, the media trunks must be protected against traditional TDM media channel attacks. TDM is a technology with already decades of history behind. Successful attacks in this plane are extremely rare.

NE Hardening: the same requirements apply as for the previous NE types.

10.5.2.4 Sentinel Gateway NE Types

NE Type (Chapter 5)	Relevant NE’s	Vulnerability (Chapter 7)	Security Patterns (Chapters 8&9)
Sentinel Gateways	A-BGF, I-BGF, FW, SBC, SEG	- Denial of service - Vehicle to attack other networks	- NE Comm. White List - Strict Protocol Adherence - NE Hardening - Topology Hiding - Communication Confidentiality - Communication Integrity - VPN

Table 10-5: Sentinel Gateway NE Types Vulnerability and Patterns

From the table we can now proceed to define specific tools and measures which derive from the fourth column.

Specific IMS Security Tools

Network Element Communication White List: although in the table above we have listed five different logical functions that can be considered to be in this class, more often than not they are seen combined into one or two network elements, which can perform all the functions. For example, it may be that a vendor sells a product which performs the function of SEG and I-BGF on the same platform. In the case of AMT, it is a matter for the operator to decide, considering cost and other advantages (scalability, geographical location of the different elements, etc.) whether it is better to separate the SEG function from the I-BGF function, and to have separate FW’s on the access side as well. If a FW element is used the white list on the outside interface will have just the FW’s IP address as its only entry. If no FW is used, there can be no white list since every valid IP public address must be accepted. The internal interface white list will be populated with the address of all elements to be reachable from the external domain.

Strict Protocol Adherence: this defense measure is specially critical for this network type, since they are the first line of defense against most attacks. Not only should it be specified during the product specification phase that these elements be extremely conservative on their acceptance of non-standardized (i.e. left up to interpretation) protocol uses, but it should also be requested that they all have the ability to report, via a management channel, any protocol abuses, so that operations personnel be aware of attempted protocol level attacks. This is also part of the responsibility of an Intrusion Detection System, as we shall see later.

NE Hardening: the same requirements as for the previous types apply here. In addition, AMT may decide based on an analysis of the hardening done on elements such as the A-BGF and I-BGF, whether a specialized FW element may be warranted, which will by definition have a more hardened OS and even hardware, than non special purpose off-the-shelf servers.

Topology Hiding: as mentioned before, this defense measure will be achieved in two ways: by how the architecture is designed, so that only two elements are facing the external domains (one for access and one for interconnect), and by the actions of the SIP layer application in those elements to remove internal information from the SIP messages. Both of these will have to be engineered correctly and jointly by the network integrator and AMT. This measure does not apply to media-only network elements.

Communication Confidentiality: with respect to signaling confidentiality, only the SBC is affected. We should remember that an SBC comprises the functions of border *control* (i.e. BCF) and border *media* (i.e. BGF). An operator may choose to deploy these two functions combined or separately. Let’s assume that AMT deploys them combined, in a single network element. In that case, the SBC will be the entity with which a user agent (the SIP subscriber terminal) establishes any security associations (i.e. encryption). So IPSec and or TLS will be requirements on this box. AMT will most likely deploy a decomposed architecture, where the border control function resides in the P-CSCF. In that case, the measure of communication confidentiality will reside there.

Communication Integrity: with respect to signaling, this measure will be included in the application of either IPSec or TLS. In the absence of either of those two protocols, integrity of signaling messages will have to be provided by the higher layers (TCP and application). UDP should be avoided for signaling.

Virtual Private Network (VPN): VPN's are the main feature of the security gateway (SEG) at the border of AMT's network and other IMS or IP networks. The SEG will be set up to establish multiple VPN's with each peering partner, engineered for the expected traffic capacity. Unlike with the security associations established in the access network (IPsec or TLS), the VPN's will transport both signaling and media to and from other networks.

10.5.3 User Scenario Specifics

In the previous section we took a service agnostic approach to defining the security requirements and architecture of AMT's IMS core. We did this by starting at the network element level, analyzing its functions according to the generic NGN conceptualized in Chapter 5, looking at possible threats, applying the corresponding patterns and selecting the tools and methods which apply to IMS. Absent from this process was any consideration of the use cases which AMT plans to market and deploy. It is necessary to at least examine these use cases in some detail, as a way of: a) verifying the security design as it stands at the moment, and b) making sure that the particulars of the use cases do not present any new threats unforeseen up to now. We now examine each scenario from the security point of view. For each scenario we look at the components, how they are used (the use case), and how they could be misused.

10.5.3.1 Voice over IP calling (handset and laptop pc)

This use case involves the following specific elements (apart from basic IMS core network elements)

a) 3G UMTS handset

- Phone itself (hardware and firmware)
- Operating system and client applications including SIP User Agent

b) Lap top PC

- 3G card with socket for UICC chip
- Operating system and client applications including SIP User Agent

The basic use case is:

1. user launches application on 3G phone or PC (can be set to launch automatically upon power on)
2. user selects “video call” and enters destination number or SIP ID by hand or from device phone book
3. user presses “send” or “make call” key
4. IMS network establishes connection to called party

Possible misuses:

1. Cloning of the UICC card in the UMTS in order to steal services and eavesdrop into conversation. Even though there have been studies published on how to extract the cryptographic key (Ki) from the USIM application in the UICC using man-in-the-middle attacks [BAR06], these attacks are highly complex. In addition, the fact that the encryption algorithm in UMTS is published (unlike in GSM), and therefore subject to constant research into how it can be broken, discovered flaws have prompted 3GPP and phone manufacturers to make the necessary changes in the protocol, to make these attacks even more difficult. In this case AMT must verify that the handsets and UICC cards deployed contain the versions of the protocol that hinder this attack.
2. Virus infection on the handset or PC. Since the handsets and the PCs use open operating systems, and users can download other applications on to them, it is possible to infect them with viruses. These might affect just the phone or the PC, in which case AMT’s network is not affected. If on the other hand the virus attempts to seize the SIP application to initiate VoIP sessions, and if enough handsets or PC’s are inffected, this could be used to launch a DoS attack on the network.

   This attack is at first view very difficult to prevent, since it uses the same modus operandi as most virus infections via the internet. The distribution of the virus to the phones could take place via free internet distribution of apps or games for the particular phone OS (e.g. Symbian, or Windows Mobile). Once in the phone, the virus can initiate sessions at random, or in coordination with other phones (i.e. at a given date and time). None of the defenses mentioned up to now in this work would prevent, or even detect that this is taking place, since the calls are legal (i.e. the device has authenticated itself). Even though the infection would eventually be detected, via customer complaints (these calls would appear on the customer bill), or by the overloads they would cause on the IMS core, the damage to the reputation of the operator would be done. A variation of this attack would be a virus which disconnects sessions after they have been initiated by the user. An even more dangerous version is a virus which would initiate random calls to 911.

AMT would need to take two additional measures as a result of this new threat: a) device and SIP application hardening, similar to the NE hardening described previously, and b) provision of virus signature detection software in the handsets.

10.5.3.2 Amateur Reporter (1-way video with 2-way speech)

This use case involves the following specific elements (apart from basic IMS core network elements)

a) 3G UMTS handset

- Phone itself (hardware and firmware) with video camera
- Operating system and client applications including SIP User Agent

The basic use case is:

1. user launches application on 3G phone or PC (can be set to launch automatically upon power on)
2. user selects “voice call” and enters destination number or SIP ID by hand or from device phone book
3. user presses “send” or “make call” key
4. IMS network establishes connection to called party
5. after agreeing with called party, user selects “one-way-video” and presses “send” or “make call” key
6. IMS network establishes one-way-video connection to called party.

Possible misuses: only the same new threats as in the previous use case, or variations of the same are thought to be possible in this use case. No new measures necessary.

10.5.3.3 Push-to-Talk over Cellular

This use case involves the following specific elements (apart from basic IMS core network elements)

a) 3G UMTS handset

  - Phone itself (hardware and firmware)
  - Operating system and client applications including SIP User Agent. It is envisioned that the same client used for voice and video calls will be used for the push-to-talk function.

b) The PoC server in AMT's application domain

The basic use case is:

1. User launches application on 3G phone or PC (can be set to launch automatically upon power on)
2. User selects "Walkie-talkie" and enters destination number or SIP ID by hand or from device phone book. User may also select group mode and invite a number of parties to the session.
3. User presses PTT key, receives signal to speak and speaks message.
4. User releases PTT key and message is routed to PoC server and from there to selected users.

Possible misuses: from the access network, only the same new threats as in the previous use case, or variations of the same are thought to be possible in this use case. No new measures necessary. We will look at other possible threats from the application domain in the Whole-network view section.

10.5.3.4 Video telephony

This use case involves the following specific elements (apart from basic IMS core network elements)

a) 3G UMTS handset

  - Phone itself (hardware and firmware) with video camera

- Operating system and client applications including SIP User Agent

b) Lap top PC with video camera

- 3G card with socket for UICC chip
- Operating system and client applications including SIP User Agent

The basic use case is:

1. user launches application on 3G phone or PC (can be set to launch automatically upon power on)
2. user selects "video call" and enters destination number or SIP ID by hand or from device phone book
3. user presses "send" or "make call" key
4. IMS network establishes video connection to called party

Possible misuses: only the same new threats as in the previous use case, or variations of the same are thought to be possible in this use case. No new measures necessary.

10.5.3.5 Location based Personalized Information Pull

This use case involves the following specific elements (apart from basic IMS core network elements)

a) 3G UMTS handset

- Phone itself (hardware and firmware)
- Operating system and client applications including SIP User Agent

b) Lap top PC

- 3G card with socket for UICC chip
- Operating system and client applications including SIP User Agent

c) The Info server (SIP and HTTP) and L&P server/enabler in AMT's application domain

The basic use case is:

1. user launches application on 3G phone or PC (can be set to launch automatically upon power on)

2. when phone registers, either after having been turned off, or as a normal periodic re-registration, IMS core informs L&P of registration and provides access network information contained in registration
3. L&P server informs Info server of new subscriber location
4. Info server gathers new information related to subscribers profile and sends a link to it in a SIP message to subscriber.
5. Subscriber clicks on the received link, which launches web browser, which sends request to Info server for the linked information.

Possible misuses: from the access network, only the same new threats as in the previous use cases, with the added variation of a browser which can also be used by the virus. No new measures necessary. We will look at other possible threats from the application domain in the Whole-network view section.

10.5.3.6 Location based Gaming

This use case involves the following specific elements (apart from basic IMS core network elements)

a) 3G UMTS handset
   - Phone itself (hardware and firmware)
   - Operating system and client applications including SIP User Agent with gaming application

b) The game server (SIP and HTTP) and L&P server/enabler in AMT's application domain

The basic use case is:

1. user launches application on 3G phone or PC (can be set to launch automatically upon power on)
2. when phone registers, either after having been turned off, or as a normal periodic re-registration, IMS core informs L&P of registration and provides access network information contained in registration
3. L&P server informs game server of new subscriber location

4. Depending on game scenarios server may inform other users of this subscriber’s location, connect two or more users via voice or video session, send user a link to game-relevant status, request status updates from other users, etc.

Possible misuses: same new threats as in the previous use cases. No new measures necessary. We will look at other possible threats from the application domain in the Whole-network view section.

10.5.4 Whole-Network View

We have now looked at the security design of AMT’s IMS core from two perspectives: individual network element and end-user use cases. Figures 10.13 and 14 summarizes the current application of security measures.

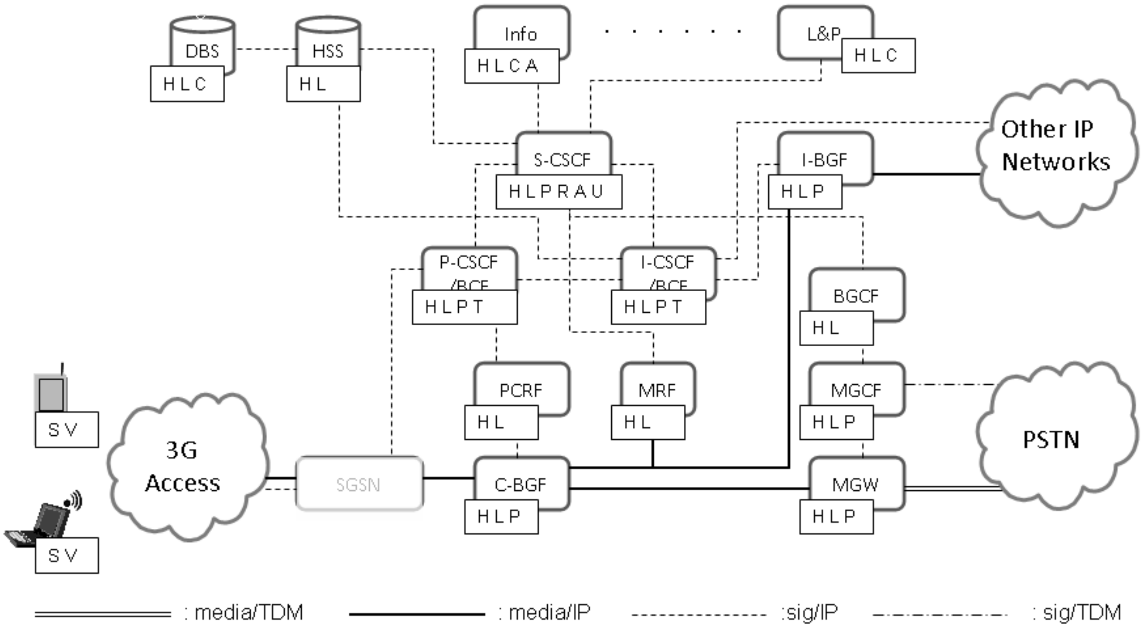


Figure 10-13 – Network Element based Security Measures (1)

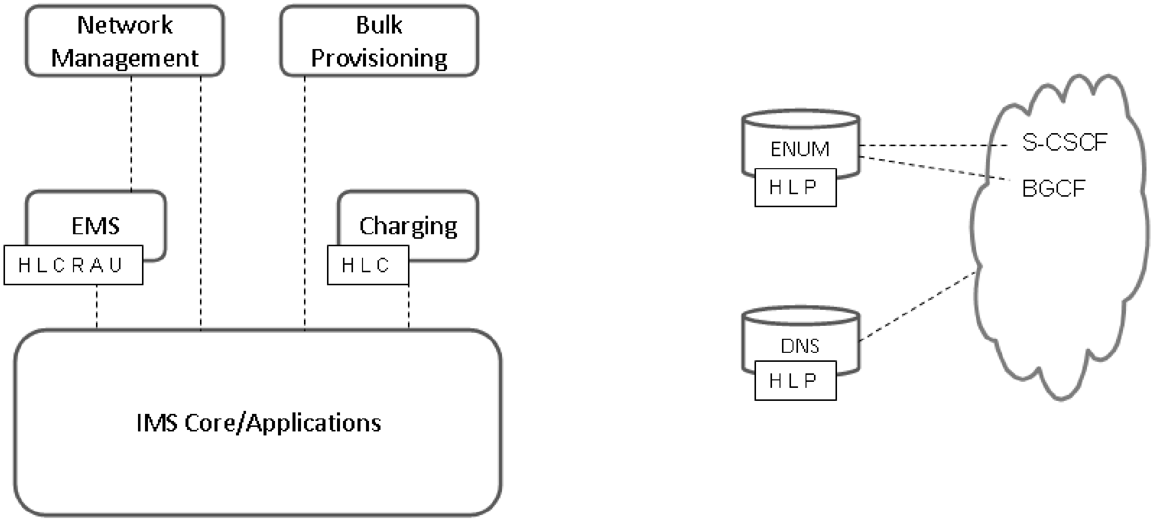


Figure 10-14 - Network Element based Security Measures (2)

In the figure, the security measures are denoted by letters to be interpreted as follows:

H : Hardening

L : White List

C : Data Confidentiality

R : Limited Registration Period

A : Authentication

U : Authorization

P : Strict Protocol Adherence

T : Topology Hiding

S : SIP Client Hardening

V : Virus Intrusion Detection

The measures of Communication Integrity (signaling), Communication Confidentiality (signaling), and non-repudiation are applied between the users and the network as shown by the solid line in Figure 10.15

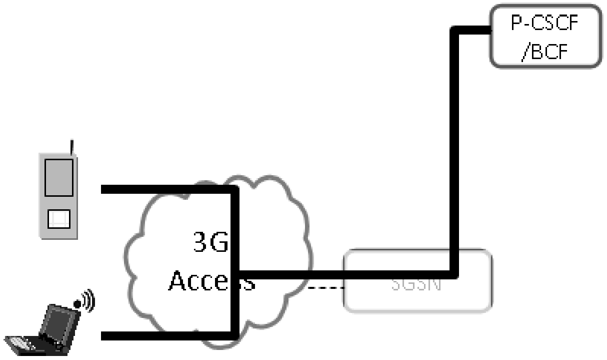


Figure 10-15 - User to Network Signaling Integrity, Confidentiality, Non-repudiation

Communications Confidentiality within IMS core network elements is recommended where shown in Figure 10.16 with the solid lines. This will always be at the discretion of the operator.

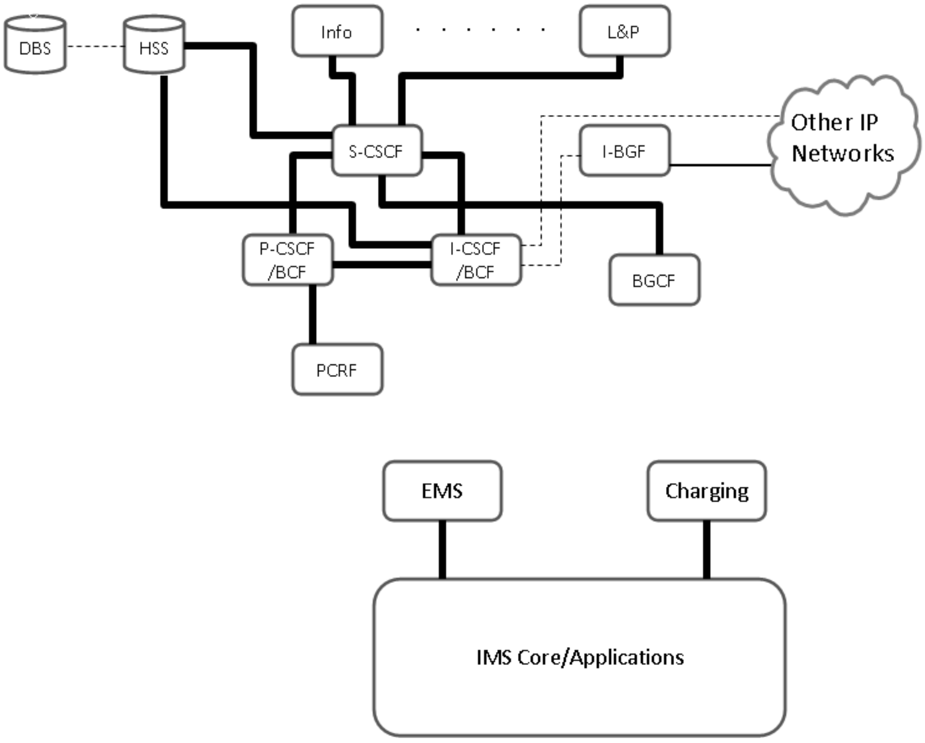


Figure 10-16 - Intra-core network Communication Confidentiality

Now, in order to finalize the security design, we need to analyze whether it makes sense to apply the rest of the security tools defined in Chapter 9. We list those here again for convenience:

a) Authentication Pattern

b) Authorization Pattern

c) Communication Confidentiality Pattern

d) Communication Integrity Pattern

e) Non-re pudiation Pattern

f) Intrusion Prevention Pattern

g) Intrusion Detection Pattern

h) Demilitarized Zone (DMZ) Pattern

i) Virtual Private Network (VPN) Pattern

j) Security Policy

The last six patterns have not been used yet. From experience with DMZ's, we know that they should be used any time an end-user is allowed to communicate directly with sensitive network components. In the case of AMT, this can happen in two cases: if users are allowed to administer and configure some of their voice features on the Voice Application Server by themselves via a web page (this is known as Subscriber Self Admin), and second, when users of the feature "Location based Personalized Information Pull" access the Info server for the personalized information.

In order to prevent these potentially risky direct accesses, AMT must provide an HTTP proxy server in a demilitarized zone. In addition, AMT should take advantage of this DMZ and also place the P-CSCF inside. The P-CSCF is the most vulnerable element in the core as it is the one to which all subscribers (hundreds of thousands or millions) have direct access to. This is shown in Figure 10.17

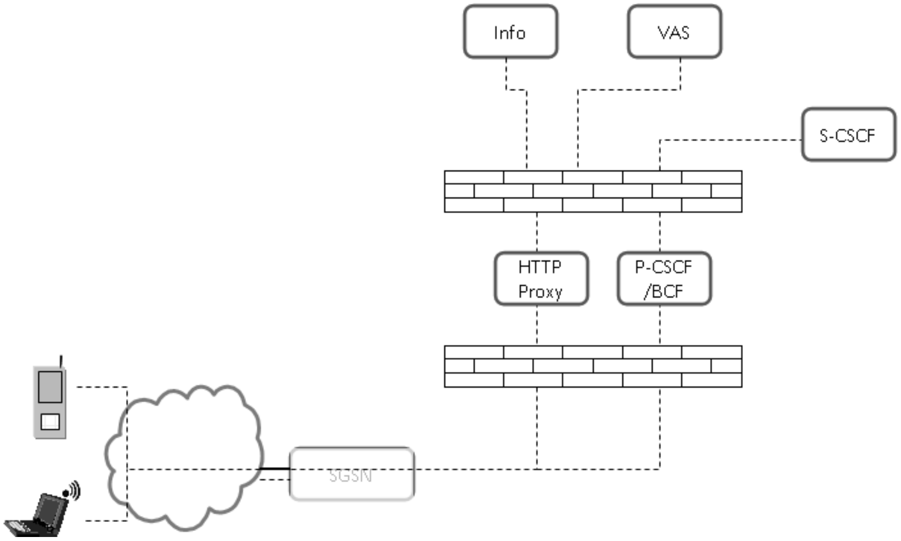


Figure 10-17 - HTTP and P-CSCF DMZ

We have also seen that one of the 3GPP interfaces, Za, towards the peering partners, requires the use of the Virtual Private Networks (VPN) pattern for both the signaling and media. This is implemented via the function Security Gateway (SEG) at the border of AMT's network. Figure 10.18 shows how this is implemented.

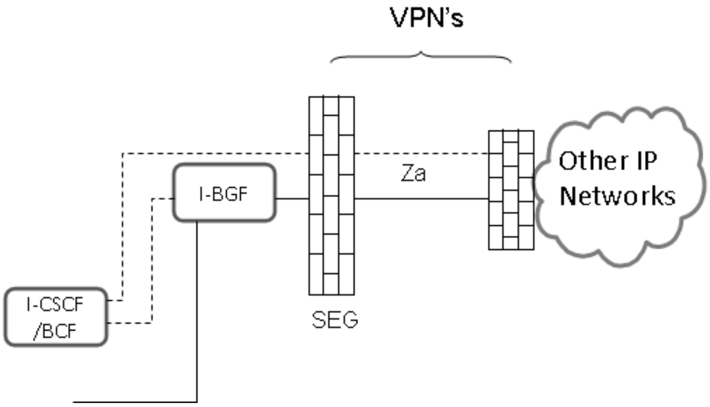


Figure 10-18 - VPN's to other IMS/IP Networks

Finally, we can provide added security protection by carefully selecting firewalls with functionality to implement the last two tools-based patterns: intrusion detection and intrusion prevention. We also provide intrusion detection capability on platforms which come in contact with the internet. Previously, in section 10.5.3.1, we had applied this to the user devices; we now apply it as well to the EMS.

An important intrusion prevention capability present in some firewalls (or Session Border Controllers) is that of being able to defend against multiple DoS attacks, especially of the malformed protocol kind. AMT needs to carefully assess the available products and choose one as the front DMZ firewall which defends against the highest number of them.

All of this is shown in Figure 10.19

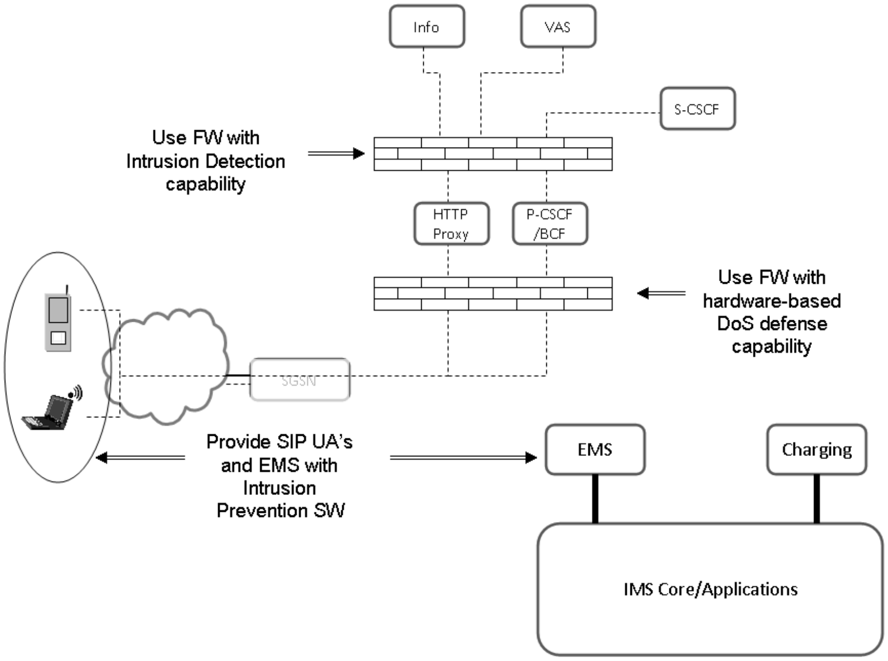

Info
VAS
S-CSCF
Use FW with Intrusion Detection capability
HTTP Proxy
P-CSCF /BCF
Use FW with hardware-based DoS defense capability
SGSN
Provide SIP UA's and EMS with Intrusion Prevention SW
EMS
Charging
IMS Core/Applications

Figure 10-19 - Intrusion Detection and Prevention, DoS Defense

The actual work of designing the security architecture for AMT's IMS core network is now finished. However, one important part of any overall security strategy still remains: security policies. We visit this in the last section of this chapter.

10.6 Security Policies

In section 9.17 the topic of security policies was introduced. Security policies are high-level guidelines for security [FER08]. Without security policies an organization of any kind does not have a mechanism with which to define, organize, and monitor its defenses and measures against likely attacks or catastrophic events.

AMT must then not only take all the steps outlined in the previous sections *prior* to the live deployment of its new IMS core network, but also define the overarching guidelines which will help it maintain a high level of security.

The definition of the security policies will usually be assigned to a specific group within the organization under the responsibility of a C-level officer, who will then have the power to enforce its provisions, and also be directly responsible for any security breaches. The security policies may apply only to the IMS core, if several parallel divisions exist within AMT, or it may apply company-wide to all the operator's networks.

What follows is a list of areas the team in charge of defining the security policies will want to consider, when drafting the security policy:

1. What should be protected

What operator and end-user property, physical or intangible, should be protected. In the case of the operator this can be obviously the actual investment in the IMS core network: hardware, software, intellectual property, sensitive information, and capital in the form of charging records. Intangible property is things such as the operator's name and reputation, contracts with other operators, and future business.

2. Rules and practices

Once the “what” has been defined, the operator needs to put in place principles, regulations, and work instructions, to help it achieve its means. The principles are the core statements from which the next two flow. They guide the decision making when conflicts or ambiguities arise. The regulations are the next level down and set about to mandate or prohibit actions and measures. Finally, the work instructions are detailed procedures which can be referred to for step-by-step directions concerning the regulations.

3. Mandated practices

Sometimes regulations come from outside the company, imposed upon it by regulatory agencies, the government, or statutory law. An example is mandates concerning information privacy, how sensitive customer information needs to be safeguarded. AMT will have to have these outside mandates into account when designing a security policy.

4. Physical access

An important part of the overall security of the network is how much access is granted to company personnel, contractors, and visitors, to sensitive physical areas. Also within certain premises, there may be equipment cabinets and racks which are unsecured, and some which are secured. A careful analysis of the area where the network equipment will be located needs to be conducted, and the rules concerning access be published and implemented, for example by implementation of electronic smart access. This analysis must also cover outside equipment (e.g. radio towers and outdoor cabinets).

5. Authorization and privilege levels

Related to number 4, this measure complements it in that it defines a Role Based Access Control method for granting privilege levels and assigns assets, areas, actions, and authority to these different levels. Operations personnel within the company (or anyone for that matter), is then assigned a level corresponding to his/her job description.

6. Obligations

Along with authorizations, obligations and responsibilities must also be assigned, so that it's clear who needs to implement policies; who need to maintain work instructions, and who can change the governing principles. The security policy must also define the personnel hierarchy, or if preferred, the role hierarchy within the company.

7. Auditing and record keeping

Any security policy must contain within itself provisions for monitoring and auditing the processes it implements. This may necessitate tools for record keeping. It must also define regular periods for review of the policy, for evaluating its effectiveness, and the methods for modifying it.

8. Disaster recovery

Finally, the security policy must have in place mechanisms for business recovery after catastrophic events whether natural, or of a terrorist nature. The results of such a policy might dictate, for example, that the IMS core network elements by duplicated in two locations, geographically separated, such that an attack on a location, or even something which affects an entire city will not render the network inoperable.

In this chapter we started with a network which had been evolved from a 2G cellular architecture which just provided voice to a 3G network designed to provide voice and multimedia applications using IMS as the control core. We then proceeded to design a security architecture using the knowledge and methodology described in the previous chapters. In the next section we draw conclusions about the task, the methods chosen, and what could be done better or researched further.

PART D

CONCLUSION

11 CONCLUSION AND OUTLOOK

In this chapter we want to summarize the thesis by reviewing what we set out to do, how we did it, what we have learned, and finally what future work could complement this study.

11.1 Conclusion

In this thesis we set out to put into practice the standards and tools developed for securing an IMS-based next generation network, by applying them to one specific case, a fictitious mobile communications operator deploying a third generation mobile network. The operator and the network are fictitious, but the architecture and applications are real and do represent many of the IMS deployments taking place today.

By undertaking this security architecture design, we expected to gain an understanding of the challenges involved, and to also come up with some methodology, a repetitive process, by which some of the uncertainty in the process and the end result could be eliminated. Needless to say that there will be more than one way in which this can be achieved, but by introducing such a method which can then be peer reviewed and critiqued, we hoped to facilitate the advancement of the techniques used for the benefit of telecommunications operators introducing IMS.

In order to be able to grasp the nature of the task at hand, we first described the players or stake holders, those who operate or make use of the networks, and their motivations, concerns, and needs. We then reviewed the state of the technology, some of its history, competing architectures, and briefly touched on the existing security standards for each. We purposely abstained from diving into the standards themselves, as this is done already in the different standards bodies and doing so would not contribute to the task at hand, but would confuse it with unnecessary detail. By describing the IMS architecture, its intended use and the needs of operators and users alike, we also set boundaries around the research domain. We expressly confined the area of study to protecting the IMS core network, keeping out domains such as the access networks, the circuit switched and packet switched GSM networks, or the IP backbone. We also listed some examples of recent publicly announced deployments to tie the theory to real on-going activities by operators all over the world.

Before concluding the review of IMS and next generation networks, in the final chapter of Part A we introduced the first tool to be used in the overall security design: the abstraction of the next generation network. The premise for this construct is the fact that by experience we observe that in any communications network, no matter what the high level functions of each of the elements, they all can be classified into discrete categories based on some basic properties. The properties chosen for the specific abstraction used here were whether the network element handled media or not, its position within the network (direct contact with un-trusted domains or not), and whether it primarily serves as a repository for data, or works on that data to accomplish an important task within the network.

In Part B, we started by examining the state of research on this topic, and verifying that no similar study has been published in the literature. We then described the nature of the threats against which the IMS network needs to be protected, including examples of the most common types of attacks, classified by their target domain. We concluded Part B by introducing some new additional tools to use in our security design, tools which we describe in the terminology of patterns. We also examined other existing patterns, which in essence, generalize all the standards and practices listed in Part A and which we purposely left unexamined there, as explained above. This, together with some specific examples of tools and protocols, completed the analysis of the threats and defenses part, and left us with a good set of building blocks to use in the methodology to be introduced in Part C.

Finally, in part C we set out to develop the process that could be used by operators or telecommunications consultants, to secure an IMS network, using the building blocks available and some knowledge about the intended use of the network. This process consists of 4 steps and is synthesized in Figure 11.1

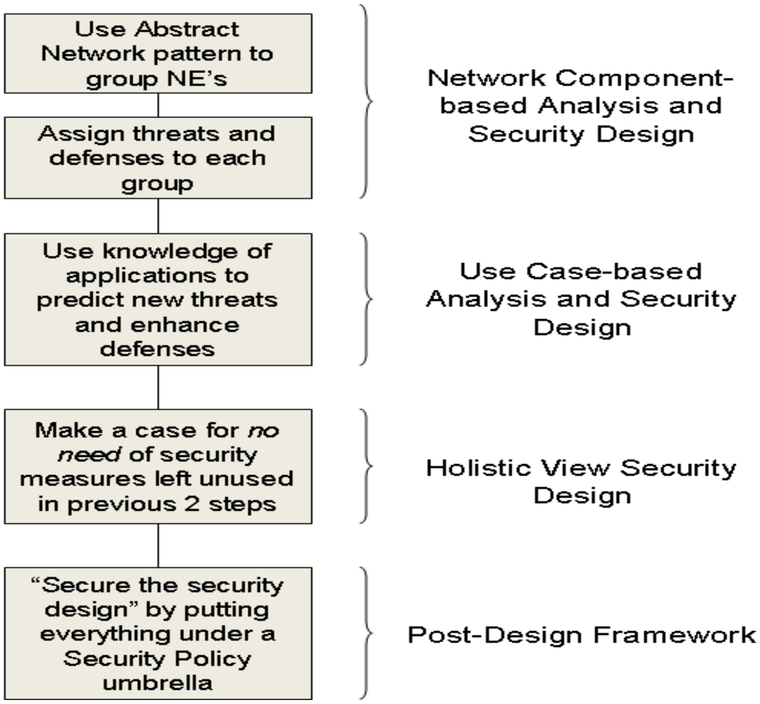

Use Abstract Network pattern to group NE's
Assign threats and defenses to each group
Use knowledge of applications to predict new threats and enhance defenses
Make a case for *no need* of security measures left unused in previous 2 steps
"Secure the security design" by putting everything under a Security Policy umbrella
Network Component-based Analysis and Security Design
Use Case-based Analysis and Security Design
Holistic View Security Design
Post-Design Framework

Figure 11-1 - Process Developed in Thesis

11.2 Lessons Learned

There is a commonly accepted truism about network security, and probably about other types of security as well, which says that security is not a feature, or a tool, or even a set of tools or hardware or software, that security is not in fact bought “off-the-shelf”. We knew going into this study that network security is an entire process that includes some or all of the above, but which is also more than that. But part of the goal of this work was to learn something additional about network security. We list below some aspects of this topic which may not have been known or emphasized prior to this study.

1. Enough tools “on the shelf”

It has been made abundantly clear that even though security is not just tools, there are more than enough of those developed and perfected, to be used in practically any type of network and useful to counter almost any kind of threat. Standard bodies like the IETF, 3GPP, TISPAN, and CableLabs have spent thousands of man-hours analyzing the threats, developing protocols, and writing documents with multiple use case examples on how to use them. These protocols and documents are publicly available for anyone to study, with the result that faults in them are usually quickly found and corrected. Software and hardware manufacturers and tool developers also have ample opportunity for turning the protocols or designs into commercially available products and freeware.

2. Customization is necessary

Notwithstanding the above, different types of operators (mobile, fixed, cable) and networks (different access types, control cores) will require that a certain degree of customization of the security solution be done. This is part of the "process" of security and cannot be avoided. For example, whether IPSec or TLS is used to protect signaling between an end-device and the control core, will depend on how the end-device access the network, directly or via network address translators. Both IPSec and TLS can be used to encrypt and to guarantee integrity, but there are use cases in which one should be used over the other.

3. Structured methodologies can be developed

Even as the number of options for defending against a given threat would seem to make for a difficult task of picking one over the other, we have shown that a methodology can be developed for systematically applying available tools, protocols, and products to a given network architecture. In this thesis we have introduced one such process, which is now available for study and optimization. As IMS is more widely deployed, others will surely be proposed.

4. New patterns can still be developed

The literature and published work on patterns is continuously expanding. It would seem that slowly the field of study would be saturated. Yet, when a particular area is made the focus of isolated research, new patterns emerge, which may have not been obvious while tackling other areas of engineering and computer networking.

11.3 Outlook

In this thesis we attempted to undertake the task of securing a next generation network based on IMS, as an operator would need to do. Although a lot of ground was covered and in the end a feasible process for doing this was developed, it is clear that not every area has been analyzed and that more detail would need to be filled in, in order for a real operator (an AT&T or a Verizon, for example) to use this work as the basis for securing their actual IMS network. Future work by the author of this thesis or by other contributors could include the following topics:

- Performance Trade-offs - How is a use case, an application, the control core itself, or the end-device affected by the introduction of a particular defense mechanism. The effect can be manifested in deteriorated quality of service as perceived by the user, or decreased computational performance of the network elements with the subsequent increase in cost to the operator.

- Reliability or robustness - When certain measures are introduced, for example the session border controller function used in our case study, a single point of failure is being put in place. If not properly architected, the element introduced to protect the network from attacks, could itself lead to total loss of service in case of internal failure.

- Recovery from breach - When any attack is successful against an IMS network, what recovery processes should be in place to return network operations to normal. Similarly, what root cause analysis and forensics tools and methods should be ready to be deployed to learn from the breach.

- Specific problems - For example, deploying TLS will entail the distribution of certificates to the servers in an IMS network, and possibly also to the end-devices. What logistical and administrative problems will this cause to the operator? Another example, what should be the re-authentication requirements when subscribers change from one access network (e.g. UMTS), to another (e.g. WiMax)? It is clear that there are many individual issues of this type which could be the target of entire papers.

This thesis has just been an introduction to the topic of *practical application* of security measures to actual operators' networks. It is hoped that future papers address some of the topics left out of this work.

12 REFERENCES

[3GP002] 3GPP TS 23.002 V7.6.0, (2008-12) Third Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Network architecture (Release 7). www.3gpp.org.

[3GP07] “3GPP Scope and Objectives”, Third Generation Partnership Project, August 31, 2007. http://www.3gpp.org/ftp/Inbox/2008_web_files/3GPP_Scopeando310807.pdf

[3GP08] “Third Generation Partnership Project Agreement”, Third Generation Partnership Project, December 4, 1988. http://www.3gpp.org/ftp/Inbox/2008_web_files/3gppagre.pdf

[3GP102] 3GPP TS 33.102, V 7.1.0 “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Security architecture (Release 7). www.3gpp.org

[3GP203] 3GPP TS 33.203 V7.5.0, 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; 3G security; Access security for IP-based services (Release 7). www.3gpp.org.

[3GP210] 3GPP TS 33.210 V7.3.0 (2007-09); “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Network Domain Security; IP network layer security (Release 7)”. www.3gpp.org

[3GP318] 3GPP TS 43.318, V 7.5.0 “Generic Access Network (GAN); Stage 2”, The Third Generation Partnership Project. http://www.3gpp.org/ftp/Specs/html-info/43318.htm

[3GP220] 3GPP TS 33.220 V7.11.0, (2008-03) “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Generic Authentication Architecture (GAA); Generic bootstrapping architecture (Release 7)”. www.3gpp.org

[3GP228] 3GPP TR 23.228 V7.14.0 (2008-12) Third Generation Partnership Project; Technical Specification Group Core Network and Terminals; IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3 (Release 7). www.3gpp.org.

[3GP229] 3GPP TR 24.229 V7.14.0 (2008-12) 3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3 (Release 7). www.3gpp.org.

[3GP310] 3GPP TS 33.310 V7.1.0 (2006-09); “3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Network Domain Security (NDS); Authentication Framework (AF) (Release 7)”. www.3gpp.org

[3GP978] 3GPP TR 33.978 V7.0.0 (2007-06) 3rd Generation Partnership Project; Technical Specification Security Aspects of early IP Multimedia Subsystem (IMS) (Release 7). www.3gpp.org.

[3GP979] 3GPP TR 23.979 V6.2.0 (2005-06) 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3GPP enablers for Open Mobile Alliance (OMA); Push-to-talk over Cellular (PoC) services; Stage 2 (Release 6). www.3gpp.org.

[3GP980] 3GPP TR 33.980 V7.6.0 (2007-09) “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Liberty Alliance and 3GPP security interworking; Interworking of Liberty Alliance Identity Federation Framework (ID-FF), Identity Web Services Framework (ID-WSF) and Generic Authentication Architecture (GAA) (Release 7)”. www.3gpp.org.

[ATT07] “AT&T moves ahead with IMS, unveils VoIP service for its IPTV customers”, Network World, January 30, 2008. http://www.networkworld.com/newsletters/converg/2008/0128converge2.html

[ATT08] ”AT&T Convergence and the Role of IMS“, AT&T White Paper available at: www.att.com/Common/merger/files/pdf/IMS_Convergence_FS.pdf

[BAR06] P. Barnard, “T-Mobile Intros New Dual Mode Wireless Service in Seattle”, IP Communications, October 2006. http://ipcommunications.tmcnet.com/hot-topics/wireless/articles/3245-t-mobile-intros-new-dual-mode-wireless-service.htm

[BAR06a] E. Barkan, E. Biham, N. Keller ” Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication”, Technion, Computer Science Technical Report, 2006-2007, Israel Institute of Technology, Haifa, Israel. www.diva-portal.org/diva/getDocument?urn_nbn_se_liu_diva-8011-1__fulltext.pdf

[BOM02] K. Boman, G. Horn, P. Howard, V. Niemi, “UMTS Security”. IEEE Electronics & Communication Engineering Journal, Volume: 14, Issue: 5, Pages(s): 191- 204, October 2002 special issue on Telecommunications Security. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1088436

[BUS96] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerland, and M. Stal., *Pattern- oriented software architecture*, Wiley 1996.

[CAB00] CableLabs, “PacketCable ™ Specifications – PacketCable 2.0”. http://www.packetcable.com/specifications/specifications20.html

[CAB00] Cable Television Laboratories, inc., (CableLabs®). www.cablelabs.org

[CAL03] Calhoun, P. et. al., Request for Comments: 3588, “Diameter Base Protocol”, September 2003, IETF Network Working Group. http://www.ietf.org/rfc/rfc3588.txt

[CAM04] Camarillo, G., Garcia-Martin, M.A., “The 3G IP Multimedia Subsystem”,John Wiley & Sons, Ltd. 2004.

[CHI08] “China Telecom selects Nokia Siemens Networks to boost its MegaEyes service for better work efficiency and life experience”, Nokia Siemens Networks, January 15, 2008. http://www.nokiasiemensnetworks.com/global/Press/Press+releases/news-archive/China_Telecom_selects_Nokia_Siemens_Networks_to_boost_its_MegaEyes_service.htm

[CHU07] “Nokia Siemens Networks signs EUR 21 million Next-Generation Network IMS contract with Taiwan’s Chunghwa Telecom” Nokia Siemens Networks, December 04,2007. http://www.nokiasiemensnetworks.com/global/Press/Press+releases/news-archive/Nokia_Siemens_Networks_signs_EUR_21_million_Next-Generation_Network_IMS.htm

[COM07] “Com hem launches first commercial IMS in Sweden”, IP TV Industry, October 24, 2007. http://www.iptv-industry.com/ar/18y.htm

[COD00] Codenomicon Defensics 3.0, Codenomicon. http://www.codenomicon.com/products/

[CTI09] “100 Wireless Facts“, CTIA the Wireless Association, 2009. http://www.ctia.org/advocacy/research/index.cfm/AID/10379

[CTI08] “U.S. Wireless Quick Facts and Figures”, CTIA The Wireless Association, May 2008. http://www.ctia.org/advocacy/research/index.cfm/AID/10323

[ETS00] “All Active Work Items for TISPAN for Current Status”, ETSI TISPAN, 2009. http://webapp.etsi.org/WorkProgram/Frame_WorkItemList.asp?SearchPage=TRUE&qETSI_STANDARD_TYPE=&qETSI_NUMBER=&qTB_ID=625%3BTISPAN&qINCLUDE_SUB_TB=True&includeNonActiveTB=FALSE&qWKI_REFERENCE=&qTITLE=TISPAN+AND+NOT+OSA&qSCOPE=&qCURRENT_STATE_CODE=&qSTOP_FLG=N&qSTART_CURRENT_STATUS_CODE=12%3BM16&qEND_CURRENT_STATUS_CODE=12%3BM16&qFROM_MIL_DAY=&qFROM_MIL_MONTH=&qFROM_MIL_YEAR=&qTO_MIL_DAY=&qTO_MIL_MONTH=&qTO_MIL_YEAR=&qOPERATOR_TS=&qRAPTR_NAME=&qRAPTR_ORGANISATION=&qKEYWORD_BOOLEAN=OR&qKEYWORD=&qPROJECT_BOOLEAN=OR&qPROJECT_CODE=&includeSubProjectCode=FALSE&qSTF_List=&qDIRECTIVE=&qMandate_List=&qSORT=TB&qREPORT_TYPE=SUMMARY&optDisplay=100&titleType=all&butExpertSearch=++Search++

[ETS001] ETSI TS 282 001 V2.0.0 “Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional architecture”, March 2008, ETSI TISPAN. http://www.etsi.org/tispan/

[ETS007] ETSI ES 282 007: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IP Multimedia Subsystem (IMS); Functional architecture", ETSI TISPAN. http://www.etsi.org/tispan/

[ETS002] ETSI ES 282 002: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); PSTN/ISDN Emulation Sub-system (PES); Functional architecture", ETSI TISPAN. http://www.etsi.org/tispan/

[ETS012] ETSI TS 182 012: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS-based PSTN/ISDN Emulation Subsystem; Functional architecture", ETSI TISPAN. http://www.etsi.org/tispan/

[ETS028] ETSI TS 182 028: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; Dedicated subsystem for IPTV functions", ETSI TISPAN. http://www.etsi.org/tispan/

[ETS182] ETSI TS 182 012 V1.1.1 "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS-based PSTN/ISDN Emulation Subsystem; Functional architecture", ETSI TISPAN. http://www.etsi.org/tispan/

[FER06] E.B.Fernandez and G. Pernul, "Patterns for Session-Based Access Control", *Procs. of the Conference on Pattern Languages of Programs, PLoP 2006,* Portland, OR, October 2006. http://hillside.net/plop/2006/

[FER06a] E.B. Fernandez, "Security patterns", *Procs. of the Eigth International Symposium on System and Information Security - SSI´2006,Keynote talk,* Sao Jose dos Campos, Brazil, November 08-10, 2006.

[FER07] E.B.Fernandez, J.C. Pelaez, and M.M. Larrondo-Petrie, "Security patterns for voice over IP networks", *Journal of Software,* Vol. 2, No 2, August 2007, 19-29. (http://www.academypublisher.com/jsw)

[FER08] E.B.Fernandez, E. Gudes, and M. Olivier, *The design of secure systems*, under contract with Addison-Wesley, 2008

[FER08a] E.B.Fernandez, G. Pernul, and M. M. Larrondo-Petrie, "Patterns and pattern diagrams for access control", *Procs. of the 5th Int. Conf. on Trust, Privacy, and Security in Digital Systems (Trustbus'08),* Turin, Italy, Sept. 1-5, 2008. Springer LNCS 5185, 38-47

[FER09] E.B.Fernandez, E. Gudes, and M. Olivier, *The design of secure systems*, under contract with Addison-Wesley.

[HAN98] Handly, et. al., Request for Comments: 2327, ”SDP: Session Description Protocol”, April 1998, IETF Network Working Group. http://www.ietf.org/rfc/rfc2327.txt

[HAR08] C. Lejnieks,”A Generation Unplugged – Research Report”, Harris Interactive, September 2008. http://www.harrisinteractive.com/News/MediaAccess/2008/HI_TeenMobileStudy_ResearchReport.pdf

[HIL05] T. Hills, “The Role of IMS in PSTN to VoIP Migration”, Light Reading Reports, December 2, 2005. http://www.lightreading.com/document.asp?doc_id=83597

[HOF06] P. Hoffman, S. Harris, “The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force”, IETF RFC 4677, IETF Network Working Group. http://tools.ietf.org/html/rfc4677

[HUN07] M. T. Hunter, R. J. Clark, F. S. Park, “Security Issues with the IP Multimedia Subsystem”, ACM MNCNA, November 2007. www.gtisc.gatech.edu/pdf/imswp.pdf

[ISC00] “Softswitch: Leading Telecommunications Companies Form International Softswitch Consortium. Group to Focus on Accelerating IP Voice and Multimedia Applications - Industry Trend or Event”, Cambridge Telecom Report, May 17, 1999.
http://findarticles.com/p/articles/mi_m0BFP/is_1999_May_17/ai_54671211

[ITU06] “H.323 : Packet-based multimedia communications systems”, International Telecommunications Union (ITU), June 2006, http://www.itu.int/rec/T-REC-H.323/e

[KOS07] G. Kostopoulos, O. Koufopavlou, “Security Analysis of SIP Signaling during NASS-IMS Bundled Authentication”. Third International Conference on Networking and Services (ICNS’07), IEEE 2007. http://www2.computer.org/portal/web/csdl/doi/10.1109/ICNS.2007.105

[KPN07] "KPN's IMS Plans Gather Pace", Next Generation Networks, September 12, 2006. http://nextgennetworks.blogspot.com/2006/09/kpns-ims-plans-gather-pace.html

[KPN08] "KPN Selects RADCOM's Solution to Monitor Services on its IMS Network", Israel On Blog, July 28, 2008. http://www.israel-on-blog.com/kpn-selects-radcoms-solution-to-monitor-services-on-its-ims-network/

[KUM09] A. Kumar, E. Fernandez, "A Security Pattern for a Virtual Private Network". www.cse.fau.edu/~**security**/papers.php

[LIB00] Liberty Alliance Project, Specifications. http://www.projectliberty.org/liberty/specifications__1

[MEY06] B. Meyerson, "Ambitious Verizon Project to cost $18 Billion", The Seattle Times, September 2006. http://community.seattletimes.nwsource.com/archive/?date=20060928&slug=verizonfiber28

[MAC09] J. A. MacDonald, "Authentication Considerations for Mobile e-Health Applications", Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on, Volume , Issue , Jan. 30 2008-Feb. 1 2008 Page(s):64 – 67. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/4562769/4571002/04571028.pdf?temp=x

[McG08] J. McGarvey, "IMS Status Report: A Protracted Adoption". Advisory Report, Current Analysis, June 20, 2008. www.currentanalysis.com/m/ericsson/CurrentAnalysis-IMS.pdf

[MSF00] "About MSF: History of Achievments", 2008, MultiService Forum, http://www.msforum.org/about/history.shtml

[MSF03] MSF-PS-MTG-REQ-001.00-FINAL “Product Specification for the Functional Requirements of an MSF Trunking Gateway”, March 7, 2003, Multiservice Switching Forum. http://www.msforum.org/techinfo/approved/MSF-PS-MTG-REQ-001.00-FINAL.pdf

[MSF05] MSF-ARCH-002.00-FINAL, “MSF Release 2 Architecture“, January 4, 2005, Multiservice Switching Forum. http://www.msforum.org/techinfo/approved/MSF-ARCH-002.00-FINAL.pdf

[MSF06] MSF-ARCH-003.00-FINAL, “MSF Release 3 Architecture“, June 2, 2006, Multiservice Switching Forum. http://www.msforum.org/techinfo/approved/MSF-ARCH-003.00-FINAL.pdf

[NCT08] “National Cable and Telecommunications Association 2008 Industry Overview”. www.ncta.com

[NEN00] “9-1-1 Fast Facts”, The National Emergency Number Association, February 2, 2009. http://www.nena.org/pages/Content.asp?CID=144&CTID=22

[NEN05] NENA Interim VoIP Architecture for Enhanced 9-1-1 Services (i2) NENA 08-001, Issue 1 December 6, 2005, National Emergency Number Association (NENA) VoIP-Packet Technical Committee. www.nena.org.

[NEN07] NENA Functional and Interface Standards for Next Generation 9-1-1 Version 1.0 (i3) NENA 08-002, Version 1.0, December 18, 2007, National Emergency Number Association (NENA) Technical Committee Chairs. www.nena.org

[NES00] Nessus® Network Vulnerability Scanner, Tenable Network Security. http://www.nessus.org/nessus/

[NOK08] “Rich Communication Suite Initiative”, Nokia Siemens Networks Press Release, Espoo, Finland, February 07, 2008. http://www.nokiasiemensnetworks.com/global/Press/Press+releases/news-archive/Rich_Communication_Suite_Initiative.htm

[OMA01] OMA-RD-PoC-V1_0, OMA PoC Specification, Requirements Document. www.openmobilealliance.org

[PRI09] D. Priselac, M. Mikuc, “Security Risks of pre-IMS AKA Access Security Solutions”. www.ericsson.com/hr/etk/dogadjanja/mipro_2008/1227.pdf

[PEL07] J. C. Pelaez, E.B.Fernandez, M.M. Larrondo-Petrie, and C. Wieser, "Attack patterns in VoIP", *Procs. of the 14th Pattern Languages of Programs Conference (PLoP2007)*, Monticello, Illinois, USA, September 5-8, 2007.
http://hillside.net/plop/2007/index.php?nav=program

[POS08] Postel J, editor, Request for Comments: 760, “DoD Standard Internet Protocol”, January 1980, Information Sciences Institute, University of Southern California. http://www.faqs.org/rfcs/rfc760.html

[QIU07] X. Qiu, N. Zhi, X. Niu, “A Quantitative Model of IMS System”, 2007 International Conference on Computational Intelligence and Security. 2007.
http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/4415275/4415276/04415462.pdf?arnumber=4415462

[RFC4975] Request for Comments: 4975, IETF Network Working Group, “The Message Session Relay Protocol (MSRP)”, B. Campbell, Ed. September 2007. www.ietf.org

[SAT09]J. Sathyan, N. Unni, “Improved Key Management Methodology for Enhanced Media Security in IMS Networks”. ICACT 2007, February 12-14, 2007.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4195308

[SHE06]M. Sher, T. Magedanz, “Developing Network Domain Security (NDS) Model for IP Multimedia Subsystem (IMS)”. Journal of Networks, Vol. 1, No. 6, November/December 2006. www.academypubli**sher**.com/jnw/vol01/no06/jnw01061017.pdf

[SHE07]M. Sher, “Secure Service Provisioning (SSP) Framework for IP Multimedia Subsystem (IMS)”. Technischen Universität Berlin PhD Thesis, 2007. http://portal.acm.org/citation.cfm?id=1163673.1163677

[SHE09]M. Sher, S. Wu, T. Magedanz, “Security Threats and Solutions for Application Server of IP Multimedia Subsystem (IMS-AS)”. www.diadem-firewall.org/workshop06/papers/monam06-paper-28.pdf

[SMI02] C. Smith, D. Collins, “3G Wireless Networks”, McGraw-Hill Telecom, 2002

[SNO00] Snort® Open Source Intrusion Prevention and Detection, Sourcefire, inc. www.sourcefire.com

[SVE07] P. Svensson, “Comcast Blocks Some Internet Traffic”, Associated Press, October 19, 2007. http://www.msnbc.msn.com/id/21376597/

[TAY08] S. Taylor, L. Hettick, “T-Mobile Offers Landline VoIP”, Network World, February 2008. http://www.networkworld.com/newsletters/converg/2008/0225converge1.html

[TEF07]”Telefonica Selects Alcatel-Lucent for Presence-based IMS Convergent Multimedia Services”, Alcatel-Lucent press release, February 12, 2007. www.info-financiere.fr/upload/FCCNS000297_20070212.pdf

[TEL07] “TeliaSonera Selects Nokia Siemens Networks' IMS Platform for Future Services”, IPTV Industry, May 8, 2007. http://www.iptv-industry.com/ar/15b.htm

[TMC06] “T-Mobile Bans VoIP and Text Messaging“, TMC Net Bloggers, May 10, 2006. http://blog.tmcnet.com/blog/tom-keating/voip/tmobile-bans-voip-and-text-messaging.asp

[PAC00] Cable Television Laboratories, inc., PacketCable™. www.packetcable.com

[PAC01] Cable Television Laboratories, inc., PacketCable Primer. http://www.packetcable.com/primer/

[PAC08] PKT-TR-ARCH-FRM-V05-080425 “PacketCable™ 2.0 Architecture Framework Technical Report”, April 2008, Cable Television Laboratories, inc. http://www.cablelabs.com/specifications/PKT-TR-ARCH-FRM-V05-080425.pdf

[PACMM] PKT-SP-MM-I04-080522 “PacketCable™ Specification Multimedia Specification”, Cable Television Laboratories, May 28, 2008. http://www.cablelabs.com/specifications/PKT-SP-MM-I04-080522.pdf

[PACRST] PKT-SP-RSTF-I04-080710 ” PacketCable™ Residential SIP Telephony Feature Specification”, Cable Television Laboratories, July 10, 2008. http://www.cablelabs.com/specifications/PKT-SP-RSTF-I04-080710.pdf

[PAR09] F.S. Park, D. Patnaik, C. Amrutkar, M. Hunter, “A Security Evaluation of IMS Deployments”, submitted for publication. Georgia Institute of Technology Computer Engineering Department, Atlanta, USA. www.cc.gatech.edu/grads/f/fpark/docs/publications/**ims**aa08_**ims**.pdf

[RIG00] Rigney, C. et. al., Request for Comments: 2865, ” Remote Authentication Dial In User Service (RADIUS)”, June 2000, IETF Network Working Group. http://www.ietf.org/rfc/rfc2865.txt

[ROS02] Rosenberg, J. et. al., Request for Comments: 3261, “SIP: Session Initiation Protocol”, June 2002, IETF Network Working Group. http://www.ietf.org/rfc/rfc3261.txt

[ROS08] Rosenberg, J. et. al., Request for Comments: 5389, "Session Traversal Utilities for NAT (STUN)", October 2008, IETF Network Working Group. http://tools.ietf.org/html/rfc5389

[SIV00] SiVuS VoIP Vulnerability Scanner. www.vopsecurity.org

[STO03] R. Stoddard, B. Dietz, "Cable Has Winning Strategy and is ready for New Season", National Cable and Telecommunications Association", April 1, 2003. http://www.ncta.com/PublicationType/Speech/531.aspx

[TIS00] Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN). http://www.etsi.org/tispan/

[TIS08] "Terms of Reference for TISPAN WG7" (Competence centre for Security), June 24, 2008, Telecommunications and Internet Converged Services and Protocols for Advanced Networking. http://portal.etsi.org/tispan/WG7_Tor.asp

[VER09] "Verizon Wireless Fosters Global LTE Ecosystem as Verizon CTO Dick Lynch Announces Deployment Plans", J. Nelson, Media Contact, Verizon Wireless, February 18, 2009, http://news.vzw.com/news/2009/02/pr2009-02-18.html

[VOD07] "WIRE: Ericsson wins IMS contracts with Vodafone Germany and Portugal", Mobile Europe, July 17, 2007. http://www.mobileeurope.co.uk/news_wire/113085/WIRE:_Ericsson_wins_IMS_contracts_with_Vodafone_Germany_and_Portugal.html