A NOVEL ARCHITECTURE FOR MOBILITY ENABLED VIDEO CONFERENCING IN NEXT GENERATION WIRELESS NETWORKS

by

AVINASH NARAYAN

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2010

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude and regards to my supervisor Prof. Sajal K. Das for his advice, encouragement, timely suggestions and guidance towards the fulfillment of my Master's Thesis. Special thanks to Prof. Mohan Kumar and Prof. Yonghe Liu for taking invaluable time serving in my thesis committee.

I would like to extend my heartfelt thanks to all my friends and fellow students at the CReWMaN laboratory without whom my work would have been very boring. Exchanging ideas with them has really been helpful towards my research work.

My research was motivated and partially supported by the NSF Grant IIS-0326505, CNS-0721951 and CNS-0916221 and AFOSR Grant A9550-08-1-0260.The work of S. K. Das is also supported by (while serving at) the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Also, thanks to Department of Computer Science for providing assistantship and scholarship for my studies.

I would further like to acknowledge Varaha System Inc., Patrick Jordan (UTA Network Services) and Craig Beal (UTA Telecommunication Services) for their assistance in providing infrastructure to setup test bed for experiments.

Above all, I am indebted to my Mother for her constant emotional and moral support and prayers without which I would not have achieved this far.

July 19, 2010

ABSTRACT

A NOVEL ARCHITECTURE FOR MOBILITY ENABLED

VIDEO CONFERENCING IN NEXT GENERATION

WIRELESS NETWORKS

AVINASH NARAYAN, M.S.

The University of Texas at Arlington, 2010

Supervising Professor: Sajal K. Das

The Heterogeneous nature of next generation networks requires architecture and technologies that have an intrinsic support for diversity. The plethora of existing wireless access technologies will soon be connected to Internet in the wave of “Internet of Things”. We have learnt from many endeavors that no single technology exists that is ubiquitous and connects everything. In Cellular service, aside from the geographic coverage limitation, the major deficiency is lack of complete coverage inside buildings (offices, healthcare facilities, malls, and the like). Once we are inside many public buildings, cellular coverage is blocked by RF opaque walls, but may have strong Wi-Fi connectivity. If our mobility solution depends on cellular services, that mobility functionality may be lost once we go inside. Providing seamless roaming and mobility wherever we go without intervention is not an added functionality but should be an in-built feature. Though many multi-mode mobile devices and technologies with video capabilities exist for a more than a decade (with the recent release of “FaceTime” calling on Wi-Fi for popular iPhone), user needs a technology that enables seamless mobility not only in Wi-Fi or GSM but across heterogeneous networks and Internet. We propose a novel architecture named as “Call Control Network Architecture (CCNA)” to enable seamless mobility in audio/video conferencing and show how CCNA achieves seamless mobility at two levels, first at core of the Internet by using a newly added network elements called “Call Control Entity (CCE)” (CCE uses existing stateful/stateless sip proxies with newly added support for including security, mobility management and call distribution mechanisms) and second at the edge of the Internet by using existing fixed-mobile-convergence / vertical handoff techniques between heterogeneous networks. Our approach is inspired by key architectural evolution techniques proposed for next generation networks in telecommunications such as SIP, IP, MPLS, FMC, VoIP etc. and key architectural evolution techniques proposed for next generation Internet such as TRIAD, SFS, DONA, HIP etc. We setup test bed based to implement seamless mobility and measure performance characteristics (such as delay, jitter, handoff time, packet loss) for the new architecture.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

The on-going rapid advances in digital audio/video processing, networking protocols, and Internet technologies have made real-time video conferencing its way into digital homes equipped consumer electronics of high performance and quality like personal computer, digital STB and hands equipped with smart phones. The next generation of enterprise networks is also undergoing rapid changes as a number of new architectures, applications, and services are being proposed and rolled out. In general, there is a convergence of voice/telephony, video, and data into a global communications network called Internet. Many factors are involved in creating a robust network capable of delivering voice/video/data services. These factors include, but not limited to, better voice and video codecs, packetization, packet loss, packet delay, delay variation, directory services, resource integration, and reliable network architecture. Also, critical are the choices of call signaling protocols, security concerns, the ability to integrate seamlessly with existing Internet services and the need to traverse network address translator (NAT) and firewalls. Video streaming over IP network is increasingly becoming the technology of choice for a wide range of network multimedia applications [16]. Two international standards exist for video conference system, namely H.323 [1] and SIP [2] recommended by ITU-T and IEFT respectively. H.323 has higher complexity and but similar signaling architecture to the traditional telephony model. SIP has been paid attention to since it has flexible and open architecture. SIP was developed by the IEFT and has been widely applied to many fields of technology including Internet telephony, video conferencing, and instant messaging. SIP along with RTP, RTSP and SDP accomplishes complete role of multimedia communication services. RTP (Real-time

Transport Protocol) [3] is used to transport real-time data and providing QoS feedback, RTSP (Real-Time Streaming Protocol) is used for controlling delivery of streaming media, and SDP (Session Description Protocol) for describing the multimedia session. The future networks are going to be heterogeneous in nature; there will be a greater need for mobility in next generation network. Video communication over heterogeneous wireline wireless networks presents several research challenges [17].

Teleconferencing has been used for more than thirty years by businesses, governments, educational institutions and other entities to enable parties in different geographic locations to communicate with one another. Teleconferencing eliminates the need for a first party to have to travel to a distant location to communicate in person with a second party, thereby saving the first party the time and expense associated with such travel. It also saves the second party the time and expense associated with having to entertain and/or host the first party. Teleconferencing is a rapidly growing application in the area of communications, in which a group of users collaborate in an interactive procedure, such as a board meeting, a task force, a scientific discussion, or even a virtual classroom, but these systems have no mobility.
Mobility plays a large role in sharing aspects of daily life; hence, people should not be tied to their devices and/or be confined to a certain location when sharing. O'Hara et al. [4] also point out the importance of mobility and the manner in which families exploit this for sharing everyday life with mobile phones.

Current teleconferencing systems typically operate by establishing a communications link over a telephone line between two different locations. Videoconferencing is a type of teleconferencing that allows parties at the two locations to speak to and to see one another. At each one of these locations, a camera, a monitor, a microphone, and a speaker are coupled to a device that interfaces with the telephone lines, wherein the camera and the microphone are used to record the visual and aural information that is to be transmitted to the other location, and the monitor and speaker are used to convey the visual and aural information recorded at the other location. They provide point-to-point and multipoint communications. In a point-to-point conference, there are two participating sites with the ability to exchange data and share user applications, while permitting the participants to hold face-to-face meetings without leaving their location. Specifically, such systems provide parties with the ability to communicate between at least two fixed locations such as a conference room, meeting room, etc. in multipoint communications, which has participation of three or more sites simultaneously. These teleconferencing systems suffer from several drawbacks. First, the production quality of such communications is typically poor. Second, such communications typically occur at low speeds. Third, these systems are more susceptible to eavesdropping and unauthorized access. It is a desire to ensure confidentiality and authenticity when teleconferencing. Finally, and most limiting, such systems have no mobility, such that the parties at one end of a communications link between two locations are only able to view what a stationary camera at the other end of the link records.

Mobility problem has been tackled in multiple ways [5, 6, 7, and 8]. Although a lot of research has been conducted in this area most of them are network centric, device centric, and focus on homogeneous networks. SIP based mechanisms for session transfer has been proposed in the IETF [8 and 9] but its extension to next generation network has not been discussed extensively. SIP also provides a framework for conferencing but in physical realization for distributed architecture there is still a centralized server that implements the focus, conference policy server and media policy server which could obviate many middle boxes, policy servers, firewalls deployed on either region of endpoints [18].

Though “Internet of Things” promises to connect everything to internet including access networks, there are many issues such as persistence, looming crisis of Internet addressing, explosion of routing table, etc. We need an architecture that is based on data and service access rather than host and IP address centered. We differ from IMS in use of flat self-certifying names, name-based routing, security and simplicity in use of network components.

This lack of mobility is a major problem for future applications. Our focus in this thesis has been to integrate some of key techniques developed to address mobility at Internet core level (TRAID[19], DONA[57], HIP[58]) and key techniques developed to address mobility at end user network level(FMC[30], IEEE 802.21[20]), in-order to come up with a synergistic architecture to enable mobility for video/audio conferencing in next generation networks connected to Internet which has in-built mechanism for distributed handling of tasks (Control Plane, Media Plane, Audio Mixing, Video Mixing, Transcoding, Transrating), acknowledges middle boxes, policy servers found on internet, and minimizes changes to network elements by using existing telephony tools such as SIP Proxy Server, ISDN Gateway, SIP Conference servers, SIP client etc. that have been rightly positioned to be the future of telephony.

1.2 Challenges

There were several challenges during implementation including coming up with the Architecture, Techniques, and Algorithms. Our test bed setup required real time transmission of video packets posed high demand on bandwidth and quality of service (QoS). Heterogeneous nature of next generation network needed a thorough understanding of various network domains, standards and technologies. Asterisk IP PBX was setup with our University's Nortel PBX with support from Telecommunications Department, Network Services Department and Varaha Inc, over several months of trials and errors. We had to deal with many Asterisk PBX issues including an issue that it does not send video if the incoming call does not have video. For developing the application for a real handset, we faced the problem that open-source SIP clients for mobile devices were unavailable. Although open-source SIP based clients for Windows and Linux exist [10] [11], the lack of open source SIP-based clients for either Windows Mobile [12] or Symbian [13] made development of this technique on handsets difficult. Most of these projects have no support for the latest Windows Mobile platform or have outdated external libraries. There were various compatibility issues regarding the libraries that were being used. The C and C++ libraries needed for Windows mobile os and present in Visual Studio were not compatible with SIP stacks. Open source Sipdroid[14] SIP/Voip client is available for Android based phones but video calls support was recently added and lacks much documentation and support. Also there is no front facing camera in Android. For iPhone Family of mobile device front facing camera and multitasking has been added to iPhone 4 recently [15]

1.3 Contributions

- Novel concept for video and audio conferencing based on heterogeneous networks
- A novel architecture to enable mobility in Video/Audio Conferencing for next generation networks, handling control plane and media plane functions in a distributed way using existing telephony tools such as SIP proxy server, ISDN Gateway, MCU, Conference Server. Please see Table 6.1 and 6.2 for more contributions and comparisons with existing conferencing models and architectures.
- Handoff Management techniques to utilize converged networks and dynamically handover based on user mobility

More details on above contributions are discussed in chapter 6.

1.4 Organization

The rest of thesis is organized as follows. Chapter 2 explores the phenomena of fixed mobile convergence, its proliferation, technology enablers, deployment and scenarios of usage. Chapter 3 describes Asterisk PBX which is used as part of our architecture and required to understand the implementation. Chapter 4 describes handoff management which is required to understand handoff technique described in this thesis. Chapter 5 describes Video over IP which is required to understand how video transmission is different from audio transmission. Chapter 6 describes our architecture of supporting mobility in video conferencing and handoff management techniques, its implementation details etc. We conclude this thesis in Chapter 7.

CHAPTER 2

FIXED MOBILE CONVERGENCE

2.1 Overview of FMC

Fixed Mobile Convergence (FMC) is an emerging technology which aims at integration and creation of a unified communication infrastructure from fixed and wireless networks. In this converged communication infrastructure, users seamlessly move across networks and access services seamlessly using different devices.

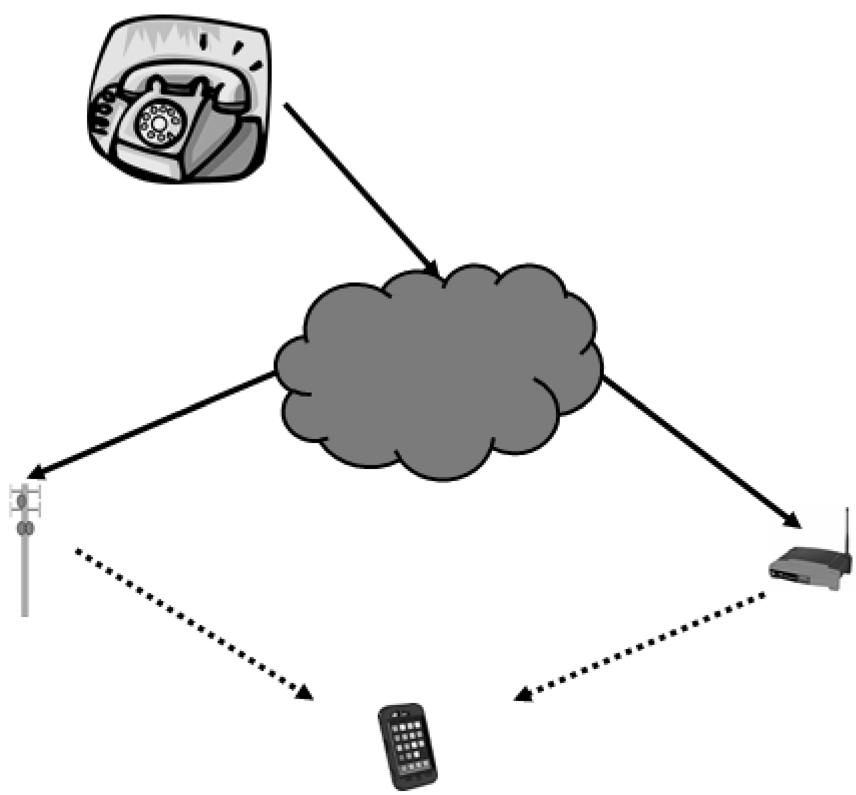

Figure 2.1 FMC with Dual Band Mobile

Voice and Video over IP (VVoIP) is one of the emerging technologies in the realization of FMC. In this chapter, we first discuss the features and technological requirements of FMC. We further identify the requirements of VVoIP and the motivation and challenges associated with migration from voice to multimedia services.

In current state where industrialized nations are quickly approaching per-capita saturation of mobile phones and the deployment of traditional landlines is quickly declining, the next generation of mobile communication has stepped onto the stage to solve problems: Fixed mobile communications, or FMC. Also referred to by other names and acronyms, FMC's core functionality is to support a cross-technology integration extending the reach of telephony beyond the limitations of cellular coverage to include support for other wireless technologies such as WiFi (IEEE 802.11), LTE [39], Mobile broadband [40] and WiMAX (IEEE 802.16) [41]. The challenge involving a cell phone roaming between two technology-compatible carriers (for example, between GSM or between CDMA) has been solved, but often accompanied by a expensive roaming charge. The latest challenge is to create new and innovative solutions that can span dissimilar wireless technologies—"to go where no mobile device has ever gone before.

The reality shows that the persons who are mobile are not the only advantage for enterprises any longer. It has become a necessary requirement for any business that wants to build solid customer relationships and take good advantage of market opportunities occurring at anytime, anywhere. The wireless devices allow both fixed and mobile employees to conduct business and carry out essential communications with their offices from remote locations, while keeping themselves updated with the latest information. But, this tendency in wireless device adoption has had some drawbacks. For many years, enterprises have built their IT environments as closed systems, protecting critical corporate information from outside intrusions. And now, when employees start communicating outside a closed environment from a remote location, such a setup causes severe limitations. The solution lies in new or re-designing a corporate infrastructure and enabling wireless communications across the enterprise [22, 23]. On other hand, mentioned above trends and business drivers in the market are pushing fixed mobile convergence (FMC) solutions enabling seamless handoff of calls across wireless and cellular networks to become a reality today. Wireline carriers are highly motivated to use FMC to reverse their loss of voice services and revenue to cellular providers. The “In-stat” reported [24] that more than 14.4% of total wireline usage has been lost to cellular, and the trend is continuing. An estimated 30% of business calls are received on cell phones even though the subscriber is in close proximity of a wireline phone and pays up to four times more for cellular calls than for wireline minutes and, in case the wireless carriers grow unconstrained, in the worst scenario, wireline carriers could risk losing their customer base entirely and becoming nothing more than wholesalers of highbandwidth pipes to wireless carriers and businesses. Also with wireless carriers racing to use wireless 3G and 4G technologies as solutions to provide last-mile voice, data, and video, they could grasp wireline carriers out of business entirely. Available today FMC solutions enable wireline carriers to recapture lost revenue and subscribers by extending mobility into the enterprise while keeping wireless minutes on the wireline network. Hence, to avoid this, wireline carriers must can utilize some of FMC solutions. These solutions use dual-mode handsets to seamlessly roam between Wi-Fi and cellular networks and provide users with one set of business telephony features, one phone number, and one user interface. This is true whatever the access mode is the enterprise Wireless LAN (WLAN), Wi-Fi hotspots, WiMAX or Cellular. FMC allows wireline carriers to retain control of the call and of their brand, turning the wireless infrastructure into a pipeline for mobile services [25, 26].

To be competent in today’s business world we need business intelligence linking and sharing in real-time. The enterprises need seamless network access as their business borders extend across local area networks (LANs) and around the world. Current mobile network connectivity includes both wired and wireless technologies. Wired connections consist of LAN ports throughout a building or campus, dial-up connectivity and secure Virtual Private Network (VPN) technology. Wireless LAN (WLAN) technology continues its increasing advancement as an essential enterprise resource. In fact, end-user spending in the WLAN enterprise-class market for WLAN access points alone is expected to increase at a compound annual growth rate of 18 % [27, 28]. We can state that the widespread adoption of WLAN has happened. Although, the barriers still (partially) exist for multimedia services such as VoIP, Video over IP, IPTV etc., IDC estimates that enterprise deployments of WLANs already exist to some degree in more than 50 % of businesses, and are expected to see substantial growth for the foreseeable future [27]. Network mobility, both wired and wireless, allows users to have access across a broad range of environments for a broad range of needs. It provides several benefits, but also brings challenges that have to be addressed with unified, holistic network architecture.

Some challenges that surround enterprise network mobility include:

- Identification and access privileges to different types of users and traffic types, with the addition of mobility provided by network
- User Bandwidth optimization, management and effective Quality of Service Control.
- Support for business applications in emerging market trends without extensive network upgrades.

Employee productivity increases with an extension of WLAN by converged network services. The studies of the “best practices” at the number of large enterprises have exposed a profound benefit: wireless mobility rapidly and positively changes the way employees work and gives them more control over their jobs. Many problems can be solved in real-time and overall collaboration is improved. This applies to not only a company’s campus workforce, but also mobile and remote employees, guests, partners and suppliers. Tied with increasing workforce and business productivity, a WLAN deployment can extend a network to areas that were previously unattached to the enterprise infrastructure, such as warehouses, stores and distribution centers. It can also boost network functionality and enable the use of new applications, such as Voice over WLAN (VoWLAN), Video over WLAN and location based services [29]

After motivating about need for Fixed Mobile Convergence here we will discuss about FMC from the point of view of Carriers/Service providers and Enterprise networks. The proliferation of broadband wireless technologies has ushered in a new era of convergence: wherein a multitude of diverse wireless devices and access technologies can be used to access a plethora of services. Such convergence, called Fixed Mobile Convergence (FMC), can be defined as convergence of wireless and wireline voice, video and broadband data services through seamless integration of wireless and fixed networks. Here, a wireless network implies a cellular network, like Global System for Mobile communications (GSM) or Universal Mobile Telecommunication System (UMTS), while a fixed network implies a Public Switched Telephone Network (PSTN) or broadband IP network. A fixed network may exist through access technologies like Digital Subscriber Line (DSL), Asymmetric DSL (ADSL) etc. In the fixed network, the last mile access network can also exist through local wireless access networks, like WiFi, Bluetooth etc., owing to their limited mobility scenario as shown in Figure 2.2.

One of the major fallouts of the FMC technology has been the creation of quad plus services in addition to quad-play services [30]. This involves FMC voice and video services with a single identity reach, voice and video call continuity, seamless mobility across access networks and devices, integrated broadband data services and simplified access for owned, subscribed and rented contents with FMC content management across networks and devices. While FMC deployment can happen within the domain of an enterprise or home network environment, larger coverage can be achieved by enabling access to different services by operators across wireless data access networks, such as WiFi, WiMAX or LTE. It involves the realization and proliferation of some of the alternate modes of communication, like Voice and

Video over IP (VVoIP). The FMC offers advantages not only for the end users but also for the service providers or operators. For the end users, it enables them to maintain a single number identity, unified billing, ubiquitous and seamless connectivity and access to consolidated set of services. For the operators, FMC reduces the stress on the available spectrum and last mile access. FMC can restrict the proliferation of base stations (BS) to provide improved signal quality to end users by operators through efficient use of available public/personal wireless local access infrastructure.

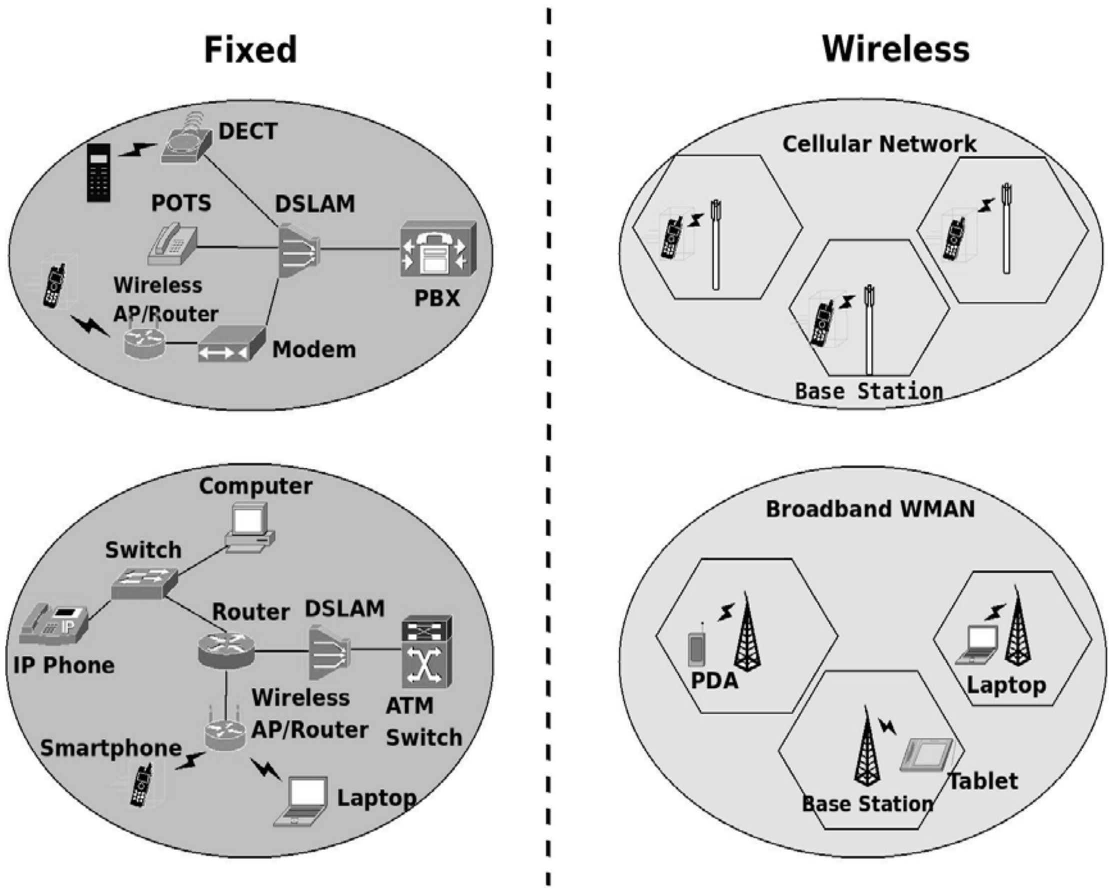

Figure 2.2 Mobility in Fixed and Wireless Networks

It enables the operators to offer user based services and gain customer loyalty. FMC aims to achieve the converged infrastructure through maximum reuse of the available technologies and mitigates the creation of yet another technology. This chapter focuses on introducing the FMC concept and discussing the primary challenges in its realization in the context of delivery of multimedia services. The rest of this chapter, we explain different types of FMC, we will consider two types of scenarios for FMC deployment and discuss the various technology enablers for realization of FMC.

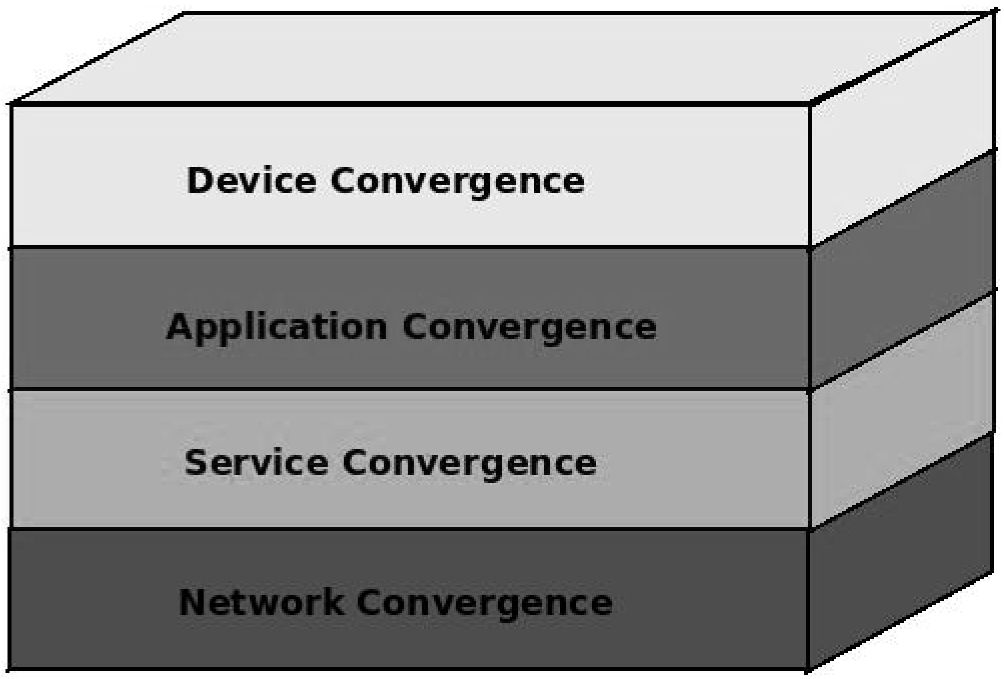

Figure 2.3 Fixed mobile convergence framework

2.2 Different types of convergence

Traditional networks architectures have been built to support specific type of service, primarily voice. To achieve FMC, a multi-level approach is required based on a variety of participating technologies. As shown in Figure 2.3, the concept of FMC revolves around the convergence of four primary levels built on each other, namely network, service, application and device [31]-[34]. They are described below.

A. Network Convergence

Network Convergence can be achieved both in the core and access networks. In the core, it implies the interoperability between the different networks types, such as IP, PSTN and Cellular, allowing seamless migration of calls and sessions between them. In access networks it primarily implies unified service delivery, continuity across access media, seamless horizontal and vertical handover across homogeneous and heterogeneous access networks, radio resource management, admission control, mobility management and flow management. To sum up, network convergence is responsible for Operations, Administration, Management and Provisioning (OAM&P) of different network components.

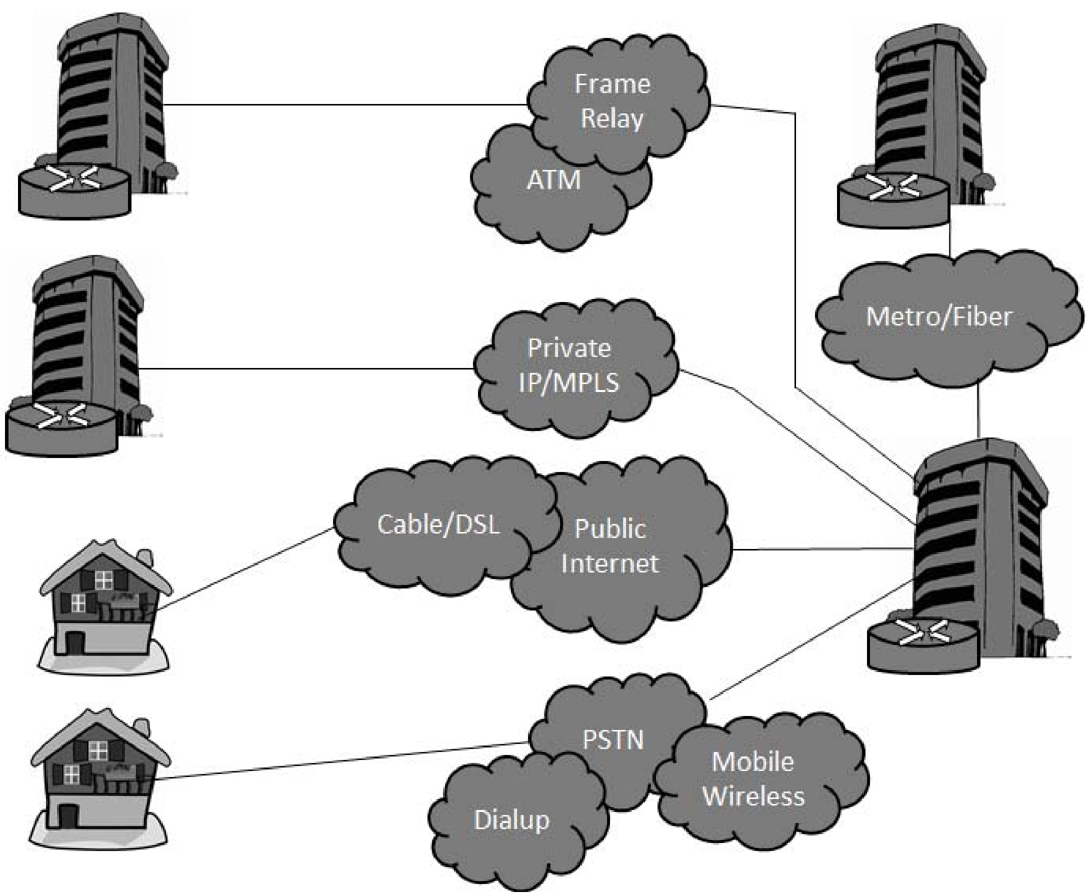

Figure 2.4 Network Convergence

B. Service Convergence

Service Convergence implies creation of a generic service framework independent of the underlying networks to enable transparent offering of services across heterogenous networks. Its functionalities include session awareness and continuity, user awareness across services, framework for deployment of user location independent policies, enable seamless roaming of terminals between wireless and wireline network domains. It requires the use of some underlying transport technologies, like Mobile IP (MIP), Unlicensed Mobile Access (UMA)/General Access network (GAN) and IP Multimedia Subsystem (IMS) to achieve these goals.

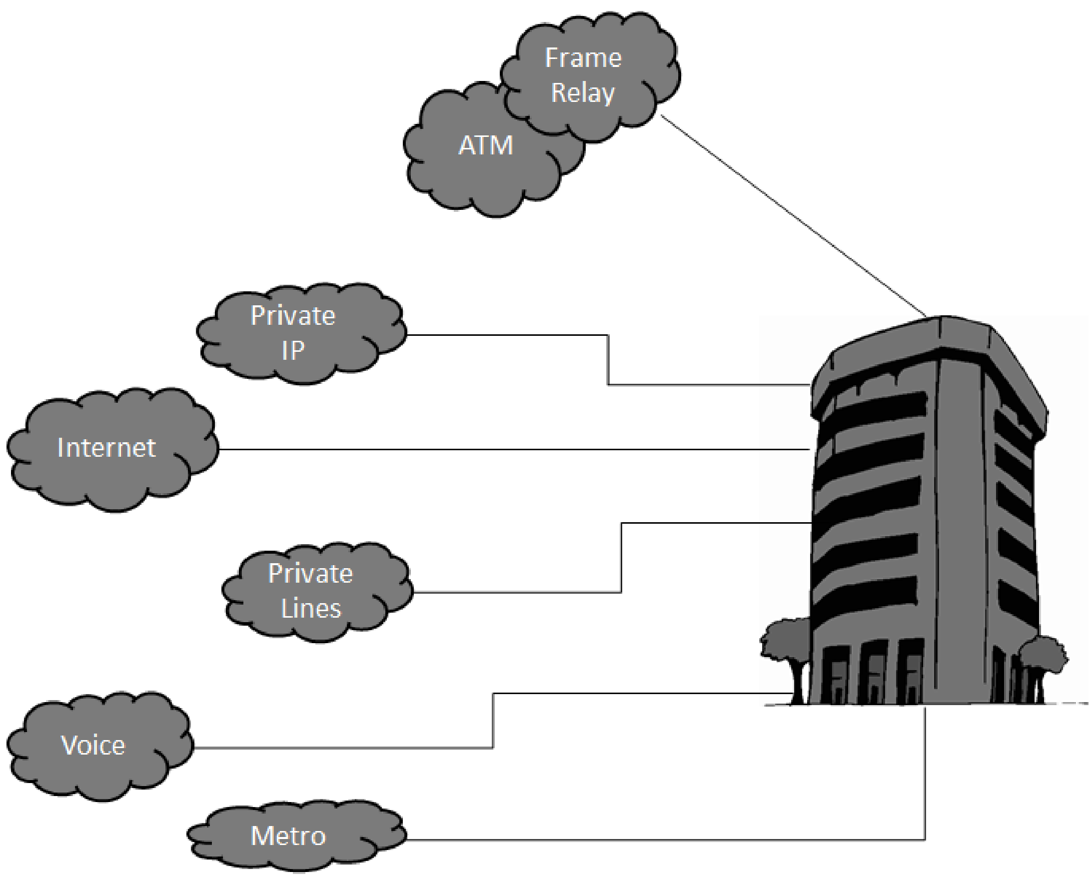

Figure 2.5 Service Convergence

C. Application Convergence

Application Convergence implies making the user experience transparent of the underlying technology. The Application layer provides the interface to the user for accessing FMC services. The clients running on end user devices provide feedback on user experience and mobility to assist network and service convergence and implement seamless interoperability between across access media, such as voice call being seamlessly continued across WiFi and cellular networks. Its functionalities may include content delivery optimization and management, network selection as well as network access optimization and management, codec selection and adaption, etc.

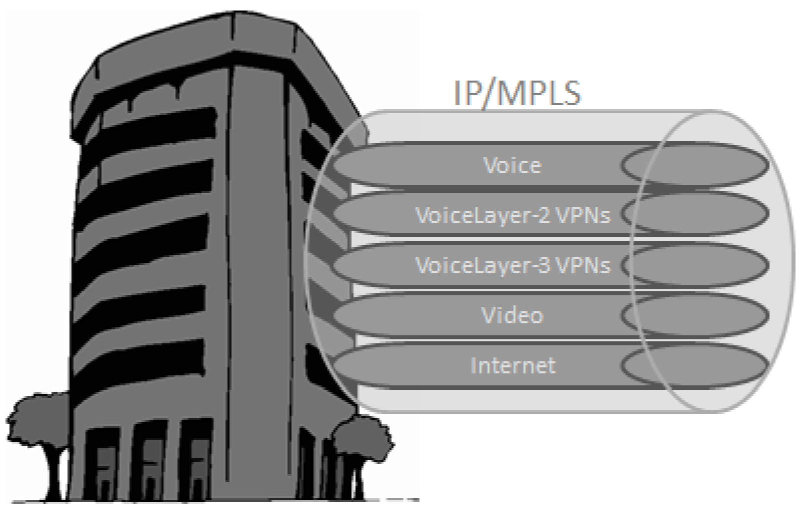

Figure 2.6 Application Convergence

D. Device Convergence

The devices used for fixed and wireless networks access have been designed on different principles. Some of the design principles, requirements and functionalities for fixed network access devices may be obsolete for wireless network access devices and vice versa. For example mobile access devices are designed for portability and battery efficiency, having advance antenna system for wide area access and accessories like camera, external storage, etc. While fixed network access devices have comparatively less complex antennas designed for small area access, not limited by battery power, inexpensive, different user interface (UI) etc. The primary requirement to achieve device convergence is to create multimodal and multifunctionality devices for ubiquitous access, having a common UI framework, unique form factor and properties as well as adaption of available devices to support FMC services.

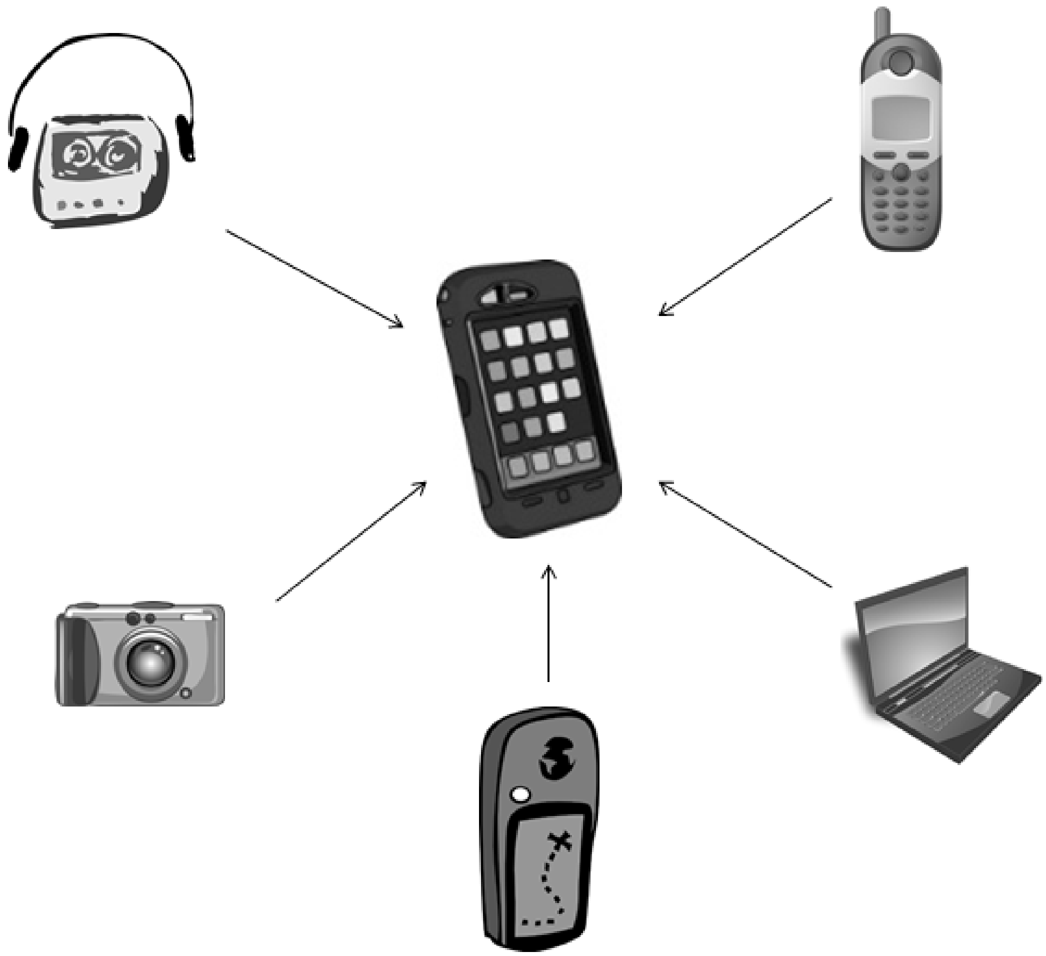

Figure 2.7 Device Convergence

FMC offers multiple advantages both from the perspective of service provider and end users. Some of the major advantages are summarized in below table.

Table 2.1 Benefits of FMC

	Benefits to End Users	Benefits to Service Providers
Cost	Maximum reuse of the available infrastructure and ubiquitous connectivity to end users at no extra cost	Operators do no need to invest in proliferation of BS to extend coverage area or improve signal quality. They can leverage the available user deployed wireless infrastructure to maximize coverage at minimum extra cost.
Connectivity and signal quality	Experiences ubiquitous connectivity and accessibility indoor as well as outdoors, since WiFi AP or femtocells are user deployed, it reduces the distance between the user terminal and them ensuring good signal quality	Can overcome dead zones indoor and outdoor through the user deployed WiFi AP or femtocells at no extra cost.
Productivity	Since last mile access is user controlled and deployed it enables him to create the network according to his mobility requirements, choice of devices to access services as well as creation, availability and deployment of service seamlessly	Enables operators to offer location and user based services as well as the independence to concentrate on outdoor coverage and mobility.
Services	*Single Number over which all services are available *Experiences integrated services ubiquitously *Can have freedom to design his own bouquet of service as well as access mechanism	Convergence enables operators to offer a single bouquet of service, both fixed and mobile as well as unified service delivery
Business Perspective	Low Cost investment required from end users	Gains customer loyalty as well as satisfaction

2.3 Fixed Mobile Convergence – Deployment Scenarios

The deployment of FMC services can be broadly classified into two scenarios - Home/Enterprise based and carrier based deployment.

A. Home/Enterprise Based FMC

There does not exist one definition for home/enterprise based FMC. It varies across service providers. In general, it can be defined as any service that enables mobile devices to communicate with and through enterprise Private Branch Exchange (PBX) over WLAN. There are three main components in the home/enterprise based FMC, namely IP/PBX, mobility controller [21], and client software, as shown in Figure 2.8, constituting the Mobile Virtual Network Operator (MVNO). Using Wireless Local Area Network (WLAN) or Femtocells, the MVNO delivers mobile as well as fixed network based services within the domain of the home/enterprise network. While the network based operations are handled by the mobility controller, the interface to the PSTN and cellular network is provided by the IP/PBX. The client software is responsible for creating a common UI framework for implementing the functionalities and access service over all networks. The MVNO is controlled and deployed by the end user. The enterprise IP/PBX can communicate with the cellular and PSTN network using a trunk gateway, such as Session Initiation Protocol (SIP), T1 or E1 trunk. By treating mobile devices as an extension of the IP/PBX, end users can be provisioned with single user identity as well as facilitate the service providers with provisioning and hosting of unified instances of value added services for the end users, e.g. voicemail, conferencing, etc. The mobility controller introduced earlier is responsible for radio resource management (RRM), mobility management, network discovery and selection, Quality of Service (QoS) control and provisioning, session continuity and redirection as well as voice call continuity etc. The mobility controller acts as a gateway for all mobile devices. The home/enterprise based FMC can be made to work with any mobile device having smartphone and multi-modal capability as well as all carriers through the IP/PBX. The client software enables to achieve device convergence through a common UI framework while is also responsible for all interaction with the mobility controller for providing assistance to enable other levels of convergence. All sessions are initiated and managed using SIP.

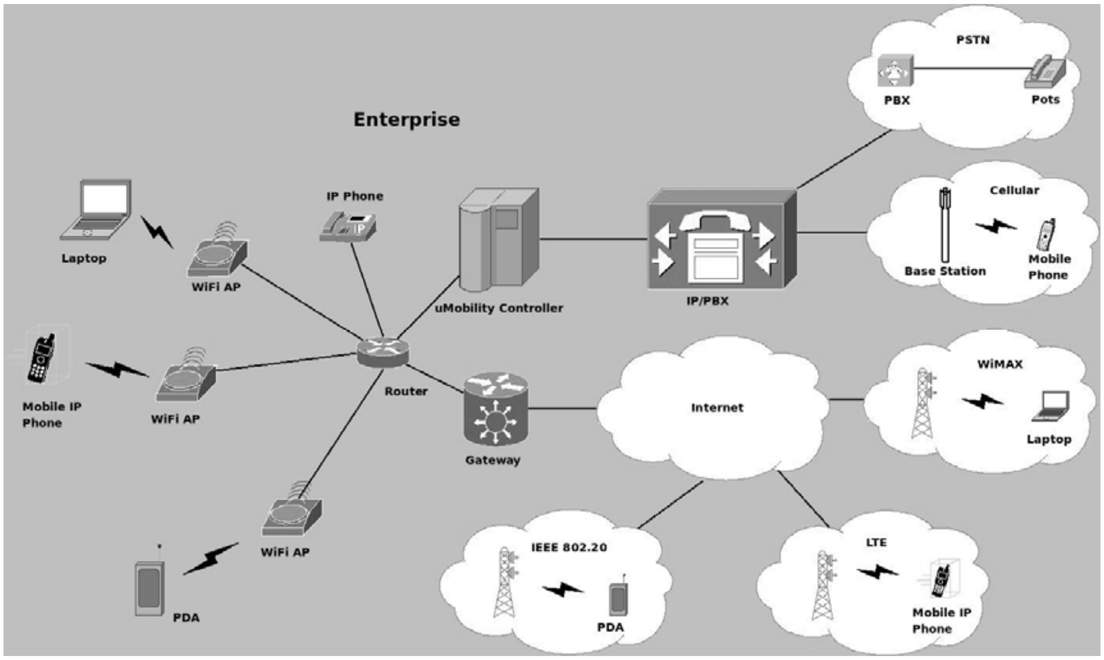

Figure 2.8 Home/Enterprise based FMC

While the home/enterprise based FMC is restricted within the domain of home/enterprise network, it offers maximum flexibility to the end user in terms of network as well as service creation and deployment, it is independent of all types of carrier core networks as well as devices and reuses the available infrastructure at minimum extra cost. With home/enterprise based FMC, the maximum control and flexibility lies with the end user.

B. Carrier Based FMC

The carrier based FMC differs from the home/enterprise based FMC in terms of the domain of the FMC services. While in home/enterprise based FMC, the coverage area is restricted to the domain of the user deployed network, in carrier based FMC it includes the whole domain of carrier network and its users as shown in Figure 2.9 below. Unlike the home/enterprise based FMC, the carrier based FMC is not carrier and device agnostic. The carrier based FMC makes use of UMA/GAN based on SIP to achieve FMC. The interface to public wireless data network is provided through cellular gateways and mobility controllers, such as GAN gateway. The disadvantage of carrier based FMC is that it is carrier specific and highly dependent on carriers for device support as well as creation and provisioning of services, which restricts the definition of FMC. It further raises various issues for operators including security, quality control, revenue impact or billing, which still need to be resolved to enable their adaptation by carriers.

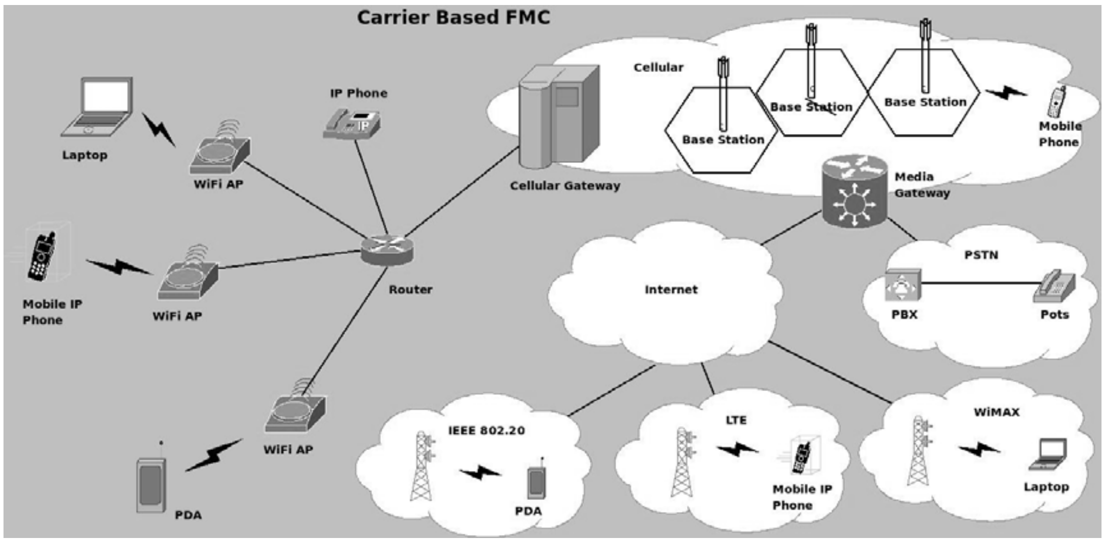

Figure 2.9 Carrier based FMC

2.4 FMC Technology Enablers

The concept of FMC is not new. Some form of FMC has existed throughout the telecommunication history. For example, when we make calls to other networks, like between cellular and PSTN network, VoIP calls to cellular or PSTN numbers etc, it is a step towards FMC. However, some of the new and emerging technologies which will help us design the next generation network (NGN) will finally help us exploit the full potential of FMC [33]-[35]. We discuss the technology enablers required for various levels of convergence.

A. Network Convergence

1) Broadband Wireless Access Technologies: The wireless access technologies can be classified into Wireless Personal Area Network (WPAN), WLAN and Wireless Metropolitan Area Network (WMAN) based on their coverage area. Among them, Bluetooth and WiFi offers attractive advantages over other wireless access technologies as they operate in unlicensed spectrum, have established support for voice, require minimum investment from the operators perspective and are ubiquitously available with low equipment cost. While research in Bluetooth is approaching saturation, some of the other WPAN technologies, like Ultra Wideband (UWB) with QoS support and data rate upto 650 Mbps offer promising future not only for voice but also multimedia applications. Similarly, the emerging 802.11n standard for WLAN promises to overcome some of the QoS restrictions of previous standards and offers high capacity with data rate up to 300Mbps to support multimedia applications. For WMAN, the emerging 802.16m standard (MobileWiMAX) and Long Term Evolution (LTE) have generated a lot of interest among operators who want to enhance the quality of service over the last mile access with high capacity.

2) Unlicensed Mobile Access (UMA) or Generic Access Network (GAN): The GAN part of the specification of Third Generation Partnership Project (3GPP) project aims at delivering mobile voice, data and IMS/SIP services over a common IP based core infrastructure. The local network can be defined by IEEE 802.11 group of standards while the WMAN is defined by UMTS based network architecture. The Generic Access Network Controller (GANC), supports RRM, mobility management, network discovery and selection etc. However this requires the use of UMA enabled handsets while the proliferation of free VoIP clients with low tariff and unlimited VoIP calls from service providers, such as Skype or Google Voice, has attracted larger user base.

3) FemtoCells: Femtocells are an emerging technology designed in view of the success of the concept of home WiFi Access Points (APs). These are low power home base stations designed to operate in home/enterprise environment, thus having low coverage area (like WiFi) but supporting WMAN access technologies (like WiMAX, LTE etc.). These base stations provide an alternate way to reap the benefits of FMC. The concept of femtocells have generated a lot of excitement in the service providers community but the onus of success of this technology lies in its ability to overcome various technological challenges and provides equipment and services at competitive rates as those currently for WiFi [36]. Thus the future of this technology is still uncertain.

4) Mobility Controller: The mobility controller [21] in the home/enterprise FMC deployment can execute various functionalities depending on its implementation and configuration. It can encompass, call admission control and continuity, session continuity and redirection, network discovery and selection, route optimization and content delivery optimization etc. In a proactive system it collects and analyzes information on the network characteristics to execute its functionality. It acts as an interface between mobile devices and home/enterprise IP/PBX or carrier core network. The mobility controller can be deployed as a in-built module of IP/PBX or act as an external module communicating with the IP/PBX using SIP.

B. Service Convergence

1) IP Multimedia Subsystem (IMS): It is an architectural framework for delivering IP Multimedia Services. It provides a common control layer which serves as an abstraction of diverse access technologies to the service layer and vice versa. It envisages the NGN to be an all IP network and thus uses SIP wherever possible. However, the IMS implementation has not been able to provide as much of cost and complexity mitigation as promised and thus has been slow in acceptance by the operators.

C. Application Convergence

1) Session Initiation Protocol (SIP): It is an application layer protocol for establishing and controlling multimedia communication sessions, like voice and video, over IP. It supports call setup and processing functionality similar to the PSTN network. Since SIP focuses only on call establishment and control, it enables us to build a variety of functionalities as in the PSTN network thus offering the same experience touser in the mobile network as in the fixed network.

D. Voice and Video over IP (VVoIP)

VoIP has emerged as a successful technology as it offers low cost, flexible, efficient network utilization as it uses IP network, user mobility as well as number portability and ease of integration with cellular and PSTN networks. The next emerging paradigm in communication technology is VVoIP, which emulates live human interaction, thus offers huge potential in revolutionizing the future modes of communication than VoIP. Advents in the last mile access technology have enabled users to experience higher bandwidth as well as emergence of IP based like SIP and IMS, are paving the way for realization of VVoIP based services. Though VVoIP offers similar technological challenges as VoIP for delivery of services over the network, it has much higher and critical constraints [37].

E. Device Convergence

1) Client Software: The home/enterprise FMC client [21] is a client software running on multi-modal devices (e.g.,smart-phones or PDAs) and provides a common UI framework for accessing various services across networks. The client is responsible for registration and interaction of the device with various functional components of the FMC in the network. The primary requirement for devices having the client software is to support dual mode operations in the network, i.e., having multiple air interfaces to connect to multiple networks simultaneously. It

is responsible for providing feedback on user experience to the mobility controller. The proliferation of smart-phones and PDAs at increasingly low cost has ensured their wide usage. The client software is generally provided by the FMC solution provider.

We also note that technologies which require wide spread adaptation by the network infrastructure have been resisted by the operators. Service Providers have been slow in adapting technologies, waiting for them to mature before deploying. Thus we need to create the FMC based network infrastructure that makes the maximum reuse of the available infrastructure. This motivates us to explore solutions based on home/enterprise based FMC for enabling global FMC access.

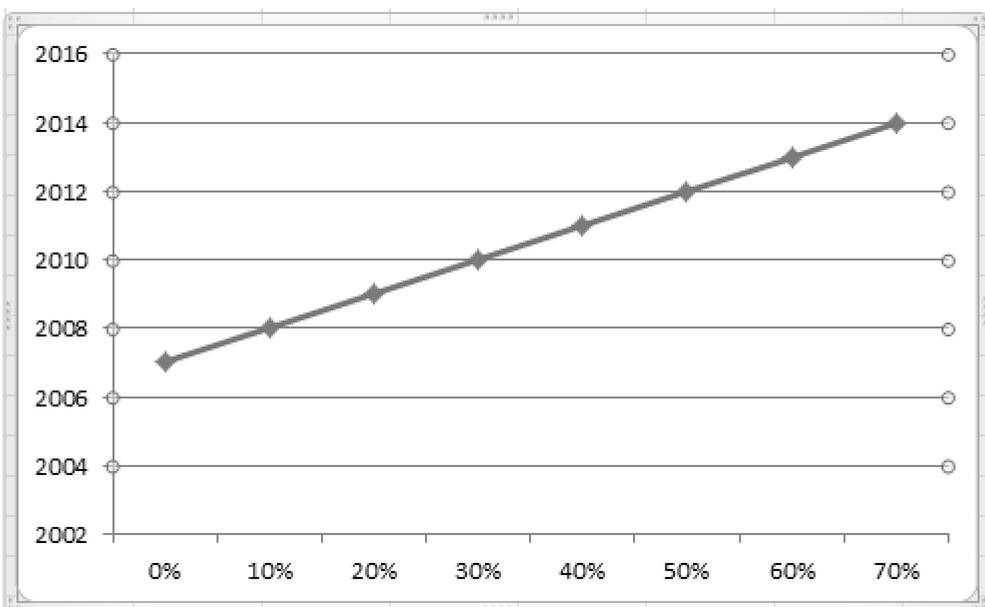

Figure 2.10 FMC penetration

The above figure shows penetration rate of convergence customers for the FMC case as noted by Renjish et al [38].

2.5 Usage Scenarios

Michael - The Marketing manager

Michael is a marketing manager in a MNC. He has a dual mode smartphone which he uses throughout an average day. Thanks to the network infrastructure and to software running on his smartphone, his device is always under company's private branch exchange (PBX) which provides necessary features, even when he is away from his desk or even not in the building. When Michael is back in the office, his smartphone connects to the internal WLAN network, and his calls are routed over the internal data network. He keeps in regular touch with the company's headquarters, and since inter-office calls are routed over the internal network, rather than the public switched telephone network (PSTN), his company cuts cost on call charges. Many advanced features are provided by his company’s PBX, including low cost internal conference calls, video conferencing, as well as document sharing and features that allow him to control how people communicate with him. For example, when he goes into a meeting, he sets his status to "meeting," and according to rules he defined, all his calls are routed to voice mail. When Michael leaves the office, his smartphone roams onto his provider's mobile network seamlessly, keeping him connected to his services with a single device. His employer is pleased because the of low cost VoIP calls that cuts down Michael's mobile phone bill, while Michael is glad because he does not have to manage multiple devices and connection lines, but all his communications happen on a single device.

Rachael – Travelling Scientist

Rachael is working as a researcher, means she has a busy travel schedule. She travels to many places to attend seminars, conferences and workshops. She has a multi-mode mobile device and a VoIP service subscription to be updated of information about her friends and family whether she is at home or on the move. Cheap Internet VoIP service allows her to save money on calls to her family and university friends that are now spread around the globe, Also her mobile operator utilizes UMA technology, which enables her to enjoy cheaper calls by using mobile phone and connecting to her home WLAN network or public hotspots. Her mobile device also enables a number of rich services which allow her to communicate with her friends via video and text as well as voice. She was able to share pictures of her travels in real-time by using mobile device and also make video message to be sent. Since all her communications are unified in a single device, Rachael's friends and family can always reach her, either by voice, text, instant message, video call or any other means, while Rachael can use presence to broadcast her availability to her contacts (such as "busy" or "travelling"), as well as manage the incoming communication depending on the context of what she is doing. The advantage is that, she need not be sitting in front of her personal computer (PC) to be able to perform many tasks that can be done on mobile. she now gets all functionalities bundled in a single device.

Camille – Home Patient

Camille is retiree and stays at home with family, she has multiple illness including diabetes, blood pressure, and arthritis. Thanks to her hospital that has deployed wireless application on a mobile device that is connected to hospital WLAN and monitors her condition. This allows doctors to review her information every day and prescribe the right medication. Camille need not visit hospital everyday hence saving money for ambulance and saves her from the strain of travelling. Even when she travels her smartphone switches automatically from WLAN to Cellular connection and she is connected all the time.

CHAPTER 3

ASTERISK: FUTURE OF TELEPHONY

3.1 Overview of Asterisk

Asterisk is an open source PBX, it has many of features available in private commercial PBX systems are including voice mail, conferencing, IVR (phone menus), and call distribution for load balancing. Much more functionality can be added by writing custom modules written in C or using Asterisk's dial plan scripts or by using Asterisk Gateway Interface Programs using any programming language capable of communicating via the standard streams system (stdin and stdout) or by network TCP sockets.

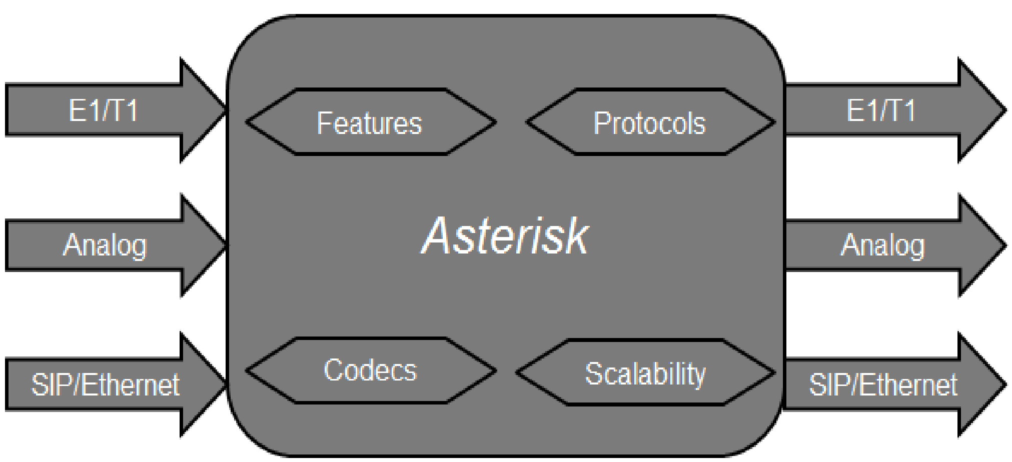

Figure 3.1 Model of Asterisk Based Exchange

We can attach traditional analogue telephones to an Asterisk installation, or to connect to PSTN trunk lines using cheap PCI Interface cards [43] manufactured by Digium and a number of other firms. These PCI cards interface with telephone lines, T1 lines, E1 lines, or any other analog/digital phone lines providing service to connected telephones/soft clients. Perhaps of more interest to many deployers today, Asterisk also supports a wide range of Video [42] and Voice over IP protocols, including SIP, MGCP and H.323. Asterisk can interoperate with most SIP telephones, acting both as registrar and as a gateway between IP phones and the PSTN. A new protocol also exists to efficiently interconnect Asterisk PBXs, called Inter-Asterisk eXchange (IAX2). By supporting a mix of traditional and VoIP telephony services, Asterisk allows to build new telephone systems, or gradually migrate existing systems to new technologies. Some sites are using Asterisk servers to replace proprietary PBX; others to provide additional features (such as voice mail or voice response menus, or virtual call shops) or to reduce costs by carrying long-distance calls over the Internet (toll bypass). Asterisk was one of the first open source PBX software packages, of which there are now many. In addition to VoIP protocols, Asterisk supports many traditional circuit-switching protocols such as ISDN and SS7 using appropriate hardware interface cards that support such protocols. Asterisk is a core component in many "PBX in a box" commercial products and Open Source projects. Some commercial products such as TrixBox and Elastix are based on Asterisk but provide additional support from manufacturers.

3.2 Configuring Asterisk

To configure Asterisk we need to edit a set of configuration files. One of these, *extensions.conf*, contains the dialplan and controls the operational flow of Asterisk. A native scripting language is used to define the elements of process control, namely named variables, procedural macros, contexts, extensions, and actions. A context groups all the valid destination numbering codes which apply to a set of channels on which incoming (to Asterisk) calls can be presented. These numbering codes, called “extensions” (even though they often are not) are the starting points for the scripts which instruct Asterisk how to process calls made to those numbers within that context. Because each channel declares a context, the dial plan restricts and permits which extensions and facilities its device may access. Extensions consist of possibly multiple steps of execution, each performing either logical operations directing program flow, or executing one of the many included applications available in Asterisk. Applications are loadable modules that perform specialized operations, such as dial a telephone number or another internal extension (*app_dial*), perform conferencing services (*app_meetme*), or handle the operations of voice mail (*app_voicemail*). These available applications provide a unique capability and tool set to formulate algorithms that can perform a large array of different, customized telephony scenarios. Applications control the Asterisk core functions through a set of internal operation primitives, that are organized in an extensible fashion through a modular architecture and application programming interfaces (APIs). Programming an Asterisk system can also be accomplished via separate, external applications using the Asterisk Gateway Interface. The Asterisk Gateway Interface (AGI) is a software interface and communications protocol for inter-process communication with Asterisk. In this, external, user-written programs, are launched from the Asterisk dial plan via pipes to control telephony operations on its associated control and voice channels. It is similar to the CGI feature of web servers in that any language can be used to write the external program which communicates with Asterisk via the standard streams, stdin and stdout.

So basically we need to perform these steps

- Define channels, contexts and user extensions based on the available technology, hardware and required number of users.
- Write a dial plan, to express the algorithm or control flow Asterisk uses to respond when calls are presented to it over these channels. Asterisk can be used for many specific applications and a customized dial plan has to be created specifically for each purpose, such as the functionality of a PBX. Asterisk is thus a 'construction kit' for building PBXs, rather than a PBX in itself, as is commonly thought.

We can also use several GUI available that makes the task of configuring asterisk easy. These interfaces allow administrators to view, edit, and change various aspects of Asterisk via a web interface. There are many GUI's available including "asterisk-gui" developed by Asterisk team and FreePBX which is usually bundled along with distributions such as AsteriskNow and TrixBox. These distributions like *AsteriskNow*, are customized Linux installation and includes FreePBX and all ancillary software to provide an "off-the-shelf" PBX, requiring only that the user prepare the requisite dial plans (see above) and connect the necessary hardware. The target market for AsteriskNow is the administrator who wishes to set up a PBX using Asterisk, but who may not have the experience in server configuration to perform the initial setup of a base Asterisk installation.

Basically Asterisk is a complete PBX in software written in C programming language and it runs on Linux operating systems. Asterisk does voice over IP in many protocols, and can interoperate with almost all standards-based telephony equipment using relatively inexpensive hardware and rivals the functionalities of traditional telephone based systems.

The benefits associated with an Asterisk based voice exchange could be summarized as:

- Cheap and inexpensive to build complex telephony systems
- Does not need a traditional PBX to begin with but can just use Ethernet
- Some features such as non-numeric extensions are only available on Asterisk
- Huge base of support from open source community
- Convergence of Voice, Video, Data on a single connection

3.3 Architecture of Asterisk Based PBX

Asterisk name says that it acts as a wildcard connecting multiple technologies and hence it has been designed for maximum flexibility. Specific APIs are defined around a central PBX core system. This advanced core handles the internal interconnection of the PBX, cleanly abstracted from the specific protocols, codecs, and hardware interfaces from the telephony applications. This allows Asterisk to use any suitable hardware and technology available now or in the future to perform its essential functions, interconnecting users, applications, hardware and protocols.

The Asterisk core handles these items internally [45]

PBX Switching, it basically interfaces with various hardware and protocols to switch calls arriving from users. It forms bridge between these and ensures the features are working as expected while switching is enabled.

Application Launcher, an application is where all the work gets done, there are hundreds of applications in asterisk including dialing number, receiving calls, IVR commands, music on hold, voice recording, scheduler etc.

Codec Translator translates calls arriving with various codec specification based on user requirements, network bandwidth, call quality etc. many codecs are supported by asterisk such as G.711, ADPCM, G.722, GSM, iLBC etc.

Scheduler and I/O Manager collaborates with host system to handle tasks related to process management, memory management, network input/output in order to keep real time traffic on application layers in constant check.

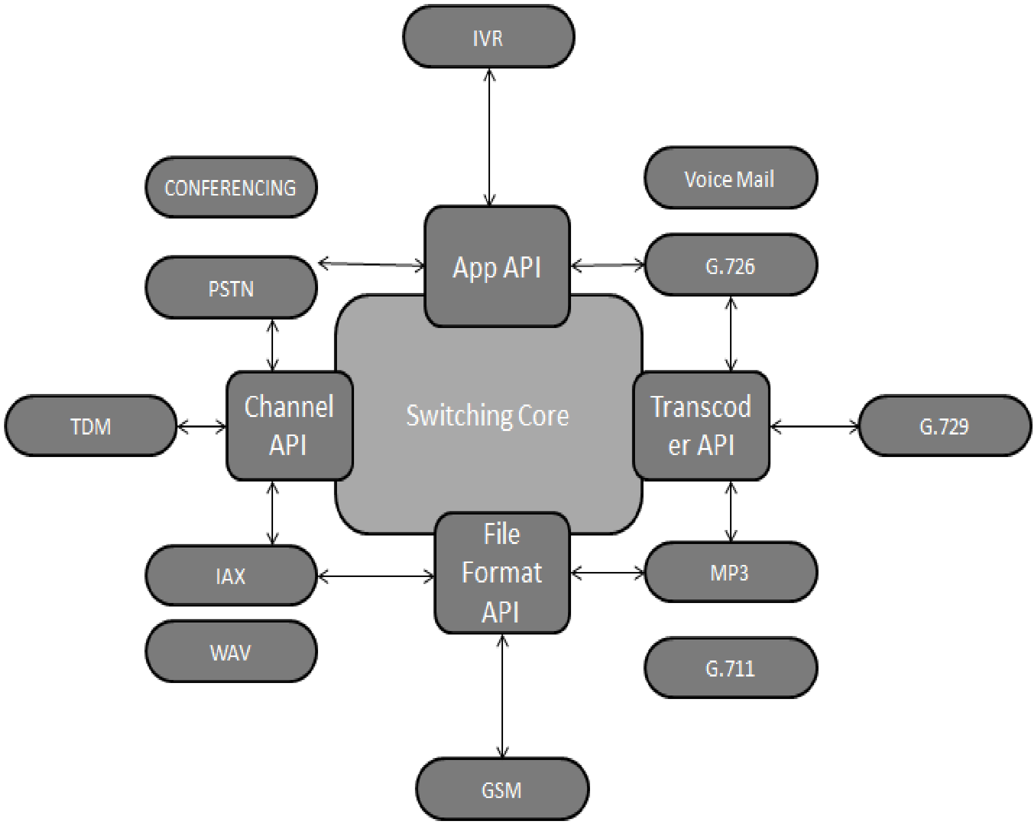

Figure 3.2 Architecture of Asterisk PBX

3.4 Asterisk Dialplan

The Dialplan is truly the core of any Asterisk system, since it defines how to handle inbound and outbound calls. It consists of a list of logical steps to follow when a call arrives. Unlike traditional phone systems, Asterisk's dialplan is flexible and customizable. To successfully set up our own Asterisk system, we will need to understand the dialplan. The dialplan is specified in a configuration file named "*extensions.conf*" and it is made up of four main concepts: contexts, extensions, priorities, and applications.

3.4.1 Contexts

(below text in brackets emphasize the syntax)

[incoming]

Dialplans are divided into sections called contexts. A context is a named groups of extensions, that can serve several purposes. Context also keeps different parts of the dialplan from interacting with one another. An extension that is defined in one context is completely isolated from extensions in any other context, unless some interaction is specifically allowed. For example, if two companies are sharing an Asterisk server. If we place each company's voice menu in its own context, they are effectively separated from each other. This allows us to independently define what happens when, say, extension 0 is dialed: people pressing 0 at Company A's voice menu will get Company A's receptionist, and callers pressing 0 at Company B's voice menu will get Company B's receptionist. Contexts are denoted by placing the name of the context inside square brackets ([ ]). The name can be made up of the letters A through Z (upper- and lowercase), the numbers 0 through 9, and the hyphen and underscore. All of the instructions placed after a context definition are part of that context, until the next context is defined. At the beginning of the dialplan, there are two special contexts named [general] and [globals]. The [general] section contains a list of general dialplan settings (which we'll probably never have to worry about), and we will discuss the [globals] context the "Global variables" section; for now it's just important to know that these two contexts are special. As long as we avoid the names [general] and [globals], we may name our contexts anything we like. When we define a channel (which is how we connect things to the system), one of the parameters that is defined in the channel definition is the context. In other words, the context is the point in the dialplan where connections from that channel will begin.

Another important use of contexts (perhaps the most important) is to provide security. By using contexts correctly, we can give certain callers access to features (such as long distance calling) that aren't made available to others. If we don't design our dialplan carefully, we may inadvertently allow others to fraudulently use our system.

3.4.2 Extensions

(below text in brackets emphasize the syntax)

exten =>

In the world of telecommunications, the word extension usually refers to a numeric identifier given to a line that rings a particular phone. In Asterisk, however, an extension is far more powerful, as it defines a unique series of steps (each step containing an application) that Asterisk will take that call through. Within each context, we can define as many (or few) extensions as required. When a particular extension is triggered (by an incoming call or by digits being dialed on a channel), Asterisk will follow the steps defined for that extension. It is the extensions, therefore, that specify what happens to calls as they make their way through the dialplan. Although extensions can certainly be used to specify phone extensions in the traditional sense (i.e., extension 153 will cause the SIP telephone set on John's desk to ring), in an Asterisk dialplan, they can be used for much more. The syntax for an extension is the word `exten`, followed by an arrow formed by the equals sign and the greater-than sign, like shown above. This is followed by the name (or number) of the extension. When dealing with traditional telephone systems, we tend to think of extensions as the numbers we would dial to make another phone ring. In Asterisk, we get a whole lot more; for example, extension names can be any combination of numbers and letters. Over the course of this chapter and the next, we'll use both numeric and alphanumeric extensions.

A complete extension is composed of three components:

- The name (or number) of the extension
- The priority (each extension can include multiple steps; the step number is called the "priority")
- The application (or command) that performs some action on the call

These three components are separated by commas, like this:

`exten =>` *`name,priority,application()`*

3.4.3 Priorities

(below text in brackets emphasize the syntax)

exten => 123,1,Answer()

exten => 123,2,Hangup()

Each extension can have multiple steps, called *priorities*. Each priority is numbered sequentially, starting with 1, and executes one specific application. As an example, the following extension would answer the phone (in priority number 1), and then hang it up (in priority num 2): The `Answer()` application is used to answer a channel that is ringing, The `Hangup()` application does exactly as its name implies: it hangs up the active channel. we should use this application at the end of a context when we want to end the current call to ensure that callers don't continue on in the dialplan in a way we might not have anticipated. The `Hangup()` application takes no arguments.

3.4.4 Unnumbered priorities

(below text in brackets emphasize the syntax)

exten => 123,1,Answer()

exten => 123,n,do something

exten => 123,n,Hangup()

Asterisk addressed the problem that needing to add something after steps have been numbered already. Which was not only difficult to change but also difficult to debug . It introduced the use of the `n` priority, which stands for "next." Each time Asterisk encounters a priority named `n`, it takes the number of the previous priority and adds 1. This makes it easier to make changes to our dialplan, as we don't have to keep renumbering all our steps. Internally, Asterisk will calculate the next priority number every time it encounters an `n`.

3.4.5 Priority labels

(below text in brackets emphasize the syntax)

exten => 123,n(label),application()

We can add labels to priorities to ensure that we can refer to a priority by something other than its number, which probably isn't known, given that dialplans now generally use unnumbered priorities.

3.4.6 Applications

Applications are the workhorses of the dialplan. Each application performs a specific action on the current channel, such as playing a sound, accepting touch-tone input, dialing a channel, hanging up the call, and so forth (*Answer(), Hangup()*). Some applications, such as `Answer()`and `Hangup()`, need no other instructions to do their jobs. Other applications require additional information. These pieces of information, called *arguments*, can be passed on to the applications to affect how they perform their actions. To pass arguments to an application, we can place them between the parentheses that follow the application name, separated by commas.

3.5 Features of SIP

Session Initiation Protocol (SIP) is an application layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. These sessions include Internet telephone calls, multimedia distribution, and multimedia [45]. SIP has the following features:

- SIP is easier to implement than available protocols such as H.323, its methods are very modular and messages have clearly defined functions.

- SIP is independent of Transport medium used to carry it, because SIP can work on UDP, TCP, ATM & so on.
- Simple text based so that it can be read by any device that can handle text processing and humans.

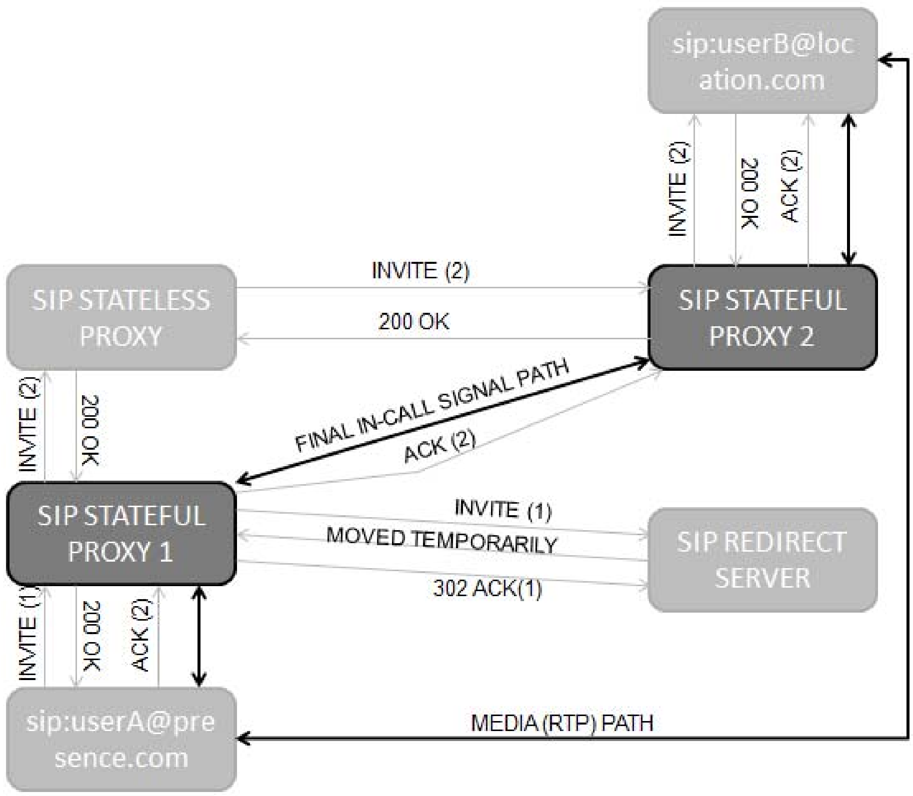

Figure 3.3 SIP Signalling

CHAPTER 4

HANDOFF MANAGEMENT

4.1 Overview of Handoff

It is the process of transferring a transparent radio link of active connections to a new channel or new service area as mobile user roams through the network system [46,47] .To provide seamless handoffs, we need to perform handoff management mechanisms that provides mobility, below we discuss some of the reasons for handoffs, types of handoffs and existing solutions

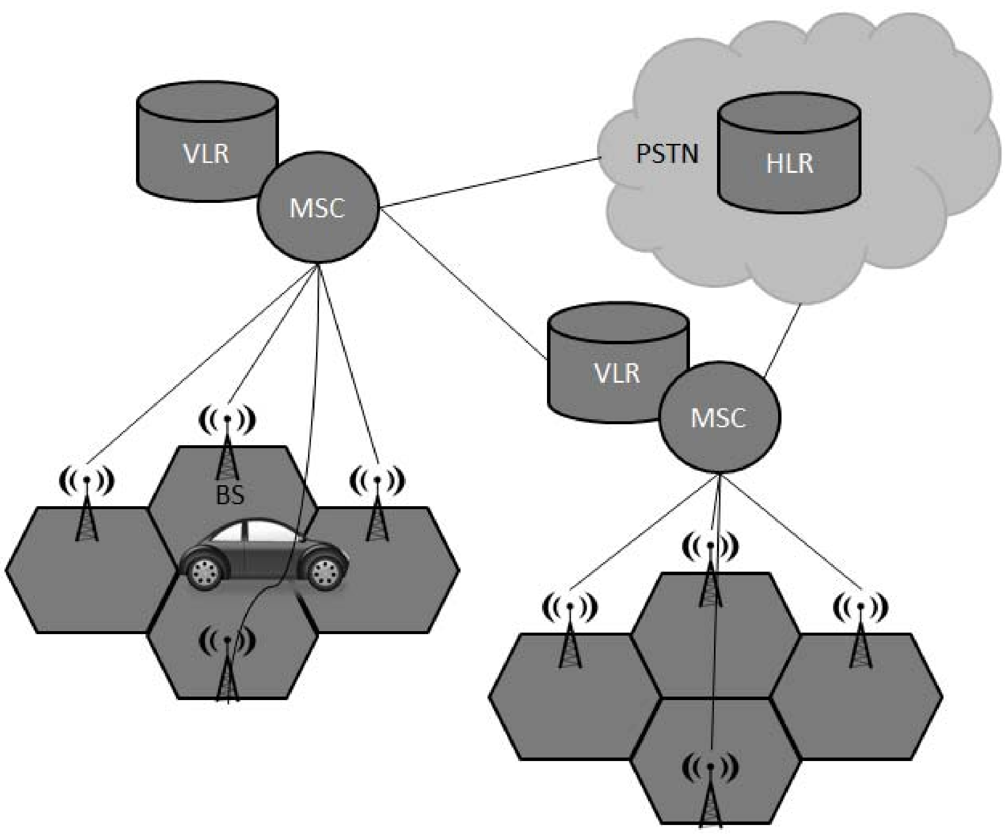

Figure 4.1 Handoffs

In telecommunications there may be different reasons why a handover might be conducted:

- When the mobile is moving away from coverage area of one cell and entering the covered area of another cell, the call is transferred to the second cell in order to avoid disconnection
- When a cell is overloaded beyond the threshold of its capacity and an existing or new call from a mobile, which is located in an area overlapped by another cell, is transferred to that cell in order to free-up some capacity in the first cell, so as to load balance and get more resources.
-in GSM networks when two mobiles are using the same channel of different cells, the call of one user is transferred to a different channel in the same cell or to a different channel in another cell in order to avoid the interference;
-in GSM networks when the mobile user behavior changes such as fast moving, inside buildings, the call is handed over to new frequency to avoid disconnection.
-in CDMA networks a handoffs can happen to improve signal quality by using softer handoffs that maintain multiple call legs with one/two base stations.

A handover, in which the target and the source are different cells is called *inter-cell* handover. The purpose of inter-cell handover is to maintain the call as the subscriber is moving out of the area covered by the source cell and entering the area of the target cell. But, in which the source and the target are one and the same cell and only the used channel is changed during the handover is called *intra-cell* handover. The purpose of intra-cell handover is to change one channel, which may be interfered, or fading with a new clearer or less fading channel. Although the idea of cellular handover or cellular handoff is straightforward, it is not an easy process to implement in reality. The cellular network needs to decide when handover or handoff is necessary, and to which cell. Also when the handover occurs it is necessary to re-route the call to the relevant base station along with changing the communication between the mobile and the base station to a new channel. All of this needs to be undertaken without any noticeable interruption to the call. The process is quite complicated, and in early systems calls were often lost if the process did not work correctly.

Many standards exist for cellular communication and each handles handover/handoff in slightly different ways. We will consider how GSM handles handover. There are a number of parameters that need to be known to determine whether a handover is required including signal strength of current base station along with the surrounding base stations. Additionally the availability of channels also needs to be known. The mobile is obviously best suited to monitor the strength of the base stations, but only the cellular network knows the status of channel availability and the network makes the decision about when the handover is to take place and to which channel of which cell.

The mobile continually monitors signal strengths of the base stations it can hear, including the one it is currently using, and it feeds this information back. When the strength of the signal from the base station that the mobile is using starts to fall to a level where action needs to be taken the cellular network looks at the reported strength of the signals from other cells reported by the mobile. It then checks for channel availability, and if one is available it informs this new cell to reserve a channel for the incoming mobile. When ready, the current base station passes the information for the new channel to the mobile, which then makes the change. Once there the mobile sends a message on the new channel to inform the network it has arrived. If this message is successfully sent and received then the network shuts down communication with the mobile on the old channel, freeing it up for other users, and all communication takes place on the new channel.

4.2 Types of handover / handoff

- Hard handover (hard handoff)
- Soft handover (soft handoff)
- Softer handover (softer handoff)
- Vertical handover (Vertical handoff)

Although all of these forms of handover or handoff enable the cellular phone to be connected to a different cell or different cell sector, they are performed in slightly different ways and are available under different conditions.

4.2.1 Hard handover

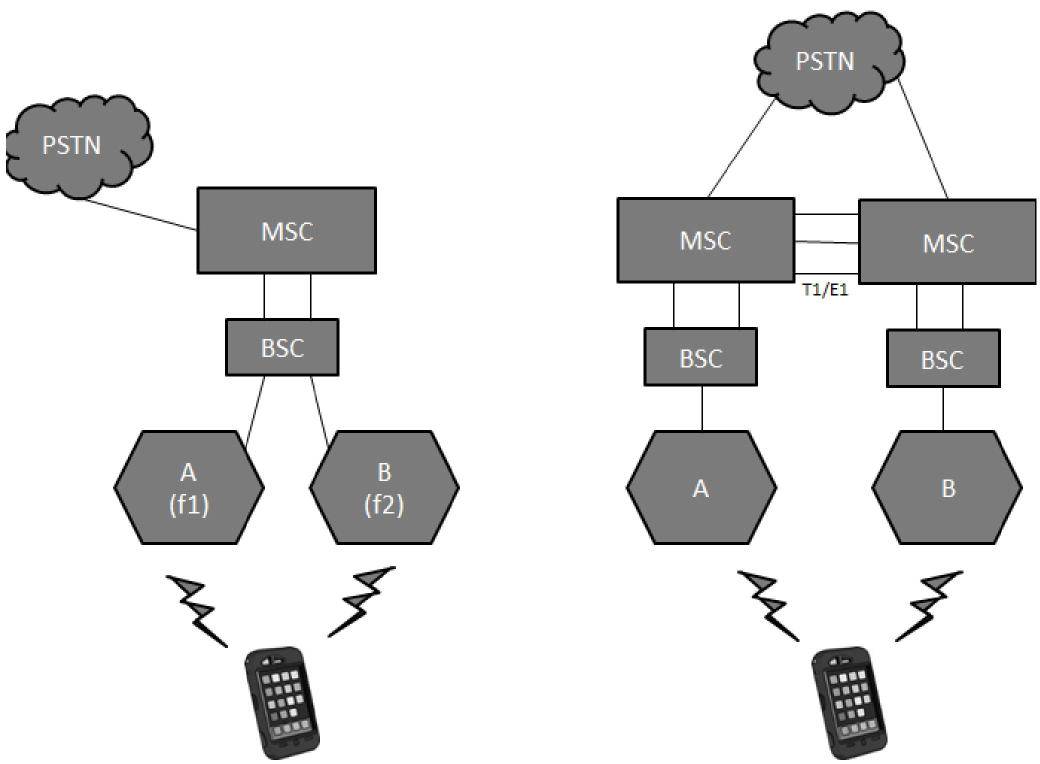

Figure 4.2 Hard Handover

Hard handover or handoff occurs when an existing connection must be broken before the new one is established. One example of hard handover is when frequencies/channels are changed. As the mobile will normally only be able to transmit on one frequency/channel at a time, the connection must be broken before it can move to the new channel where the connection is re-established. This is often termed as inter-frequency hard handover. While this is the most common form of hard handoff, it is not the only one. It is also possible to have intra-frequency hard handovers where the frequency channel remains the same. Although there is generally a short break in transmission, this is normally short enough not to be noticed by the user.

4.2.2 Soft hand over

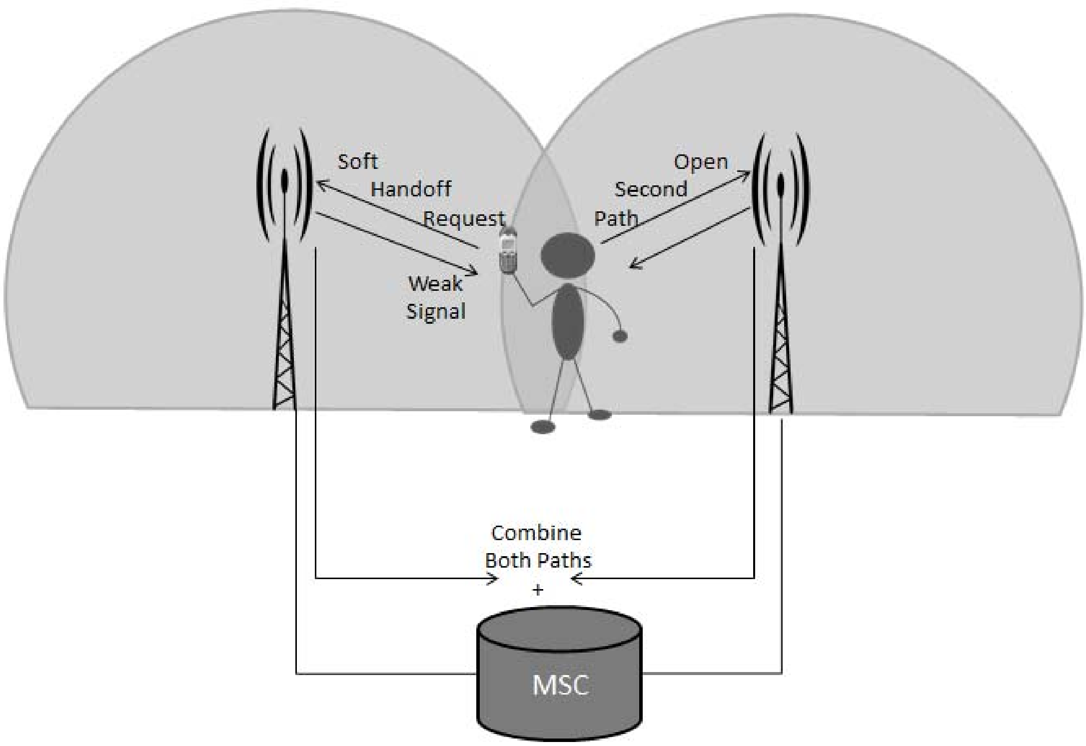

Figure 4.3 Soft Handoff

In CDMA where it is possible to have neighboring cells on the same frequency soft handover or handoff occurs where connection is established with two cells which operate on the same frequency/channel.

4.2.3 Softer handover

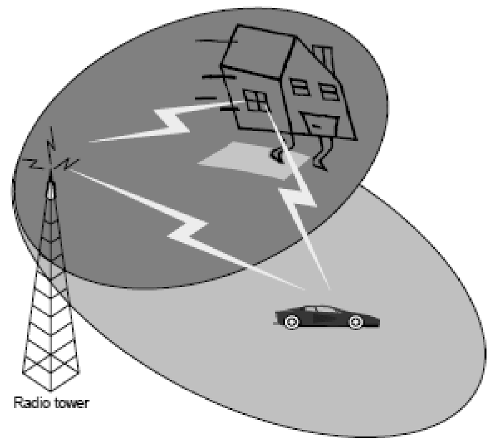

Figure 4.4 Softer Handoff

In this form of handoff, two legs are accessed on the same cell but on different sectors, which have antennas to divide cells into different sectors.

4.2.4 Vertical Handover

Vertical handoff occurs between different technologies with mobile devices enabled with multiple access technologies usually to maintain mobility. For example a dual mode mobile providing both WLAN and cellular connectivity interface can be connected to either network when the other is not available thus enabling user to be connected all the time. Vertical handoff decision in an Next Generation of network environment is more complex and involves a tradeoff among many handoff metrics including QoS requirements (such as network conditions and system performance), mobile terminal conditions, power requirements, application types, user preferences, and a price model [48], it also refer to the switching from one technology to another automatically in order to maintain communication.

4.3 Handover in FMC network

There are two types of handover in a FMC based deployment scenarios as explained below.

4.3.1 Enterprise based Seamless Handover

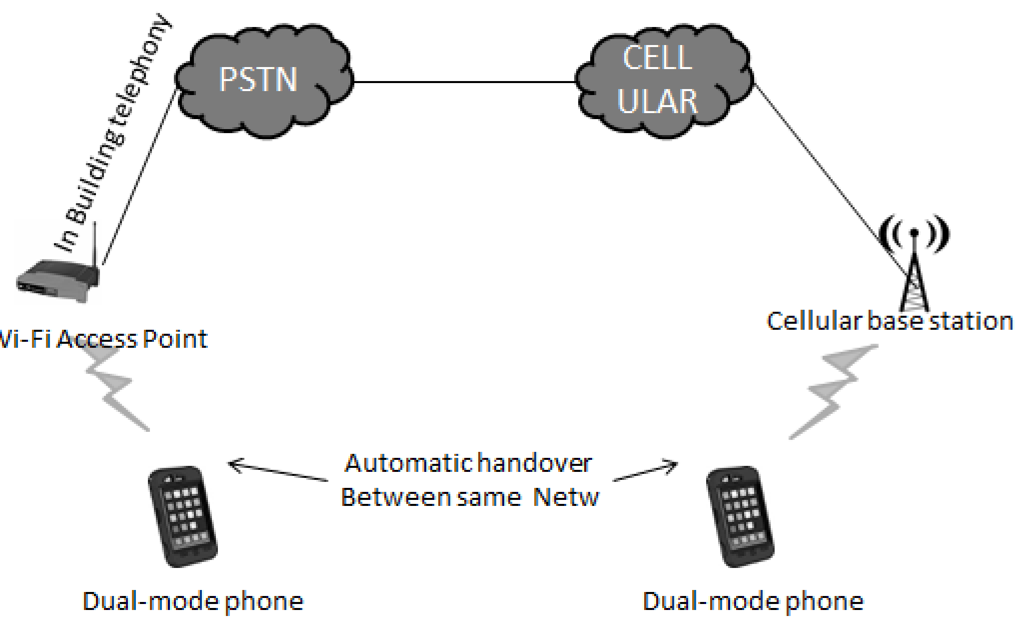

Figure 4.5 Enterprise Seamless Handover

Architecture proposed by 3GPP supports "seamless handover" service that be part of the infrastructure, requiring no user intervention, also a number of solutions in market are based on this architecture. This core feature makes no assumptions as to where the call may be anchored (PBX or cellular) but provides the user with a seamlessness not offered with the other implementation options. This is a true FMC deployment but this later approach is more complex and has more market and technology barriers to address.

4.3.2 Carrier Based Seamless Handover

In the case of a UMA solution (see Figure 4.6), the dual-mode phone uses the basic IP transport of a WiFi access point in the same manner that it would use a standard cellular mobile switching center (MSC). Carrier-based handoff and UMA and voice call continuity (VCC) provide an automatic handoff between WiFi and cellular networks but differ where the call control is resident

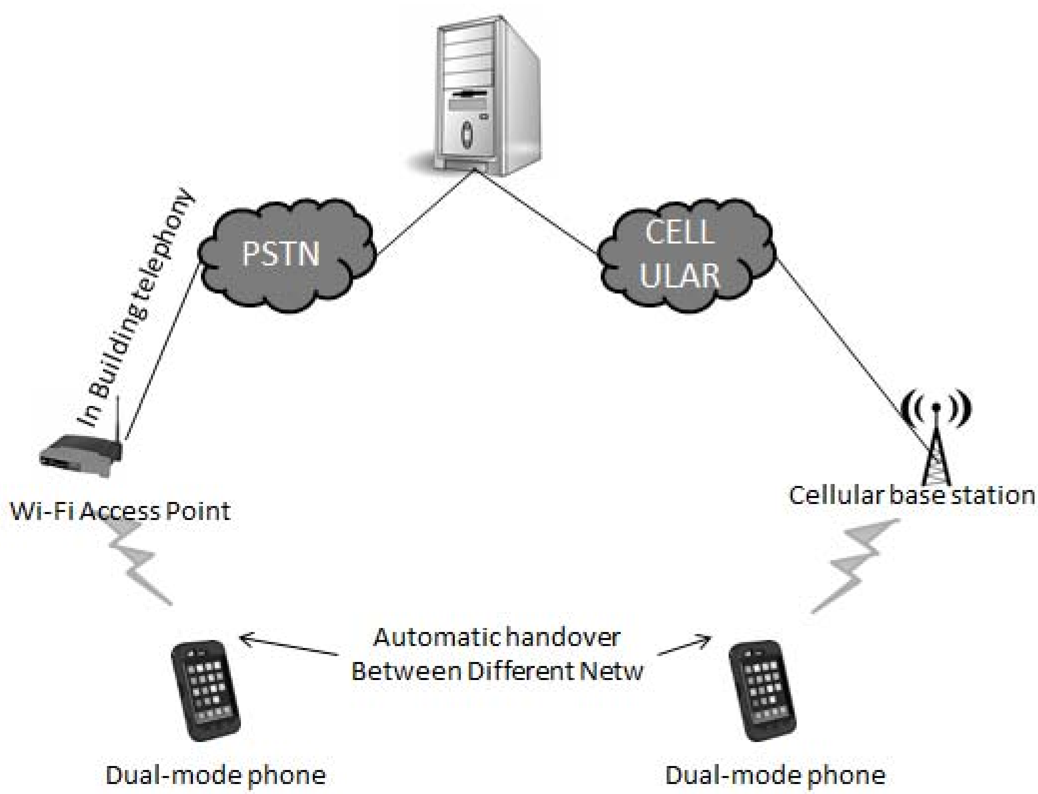

Figure 4.6 Carrier Based Seamless Handover

In this architecture a seamless handover is part of the carrier network infrastructure

4.4 Implementations

Handoffs are implemented in cellular systems using complex set of algorithms and tools that predict many changes including signal strength deterioration, network load, and quality of service. Handoffs are done based on creating a neighbor list. This list is used by both mobile device and cellular system to create a transfer path for calls from one cell to another

Cellular system monitors many parameters in the source cell and assessed in order to decide when a handover may be necessary. The downlink (forward link) and/or uplink (reverse link) directions may be monitored. The handover may be requested by the phone or by the base station (BTS) of its source cell and, in some systems, by a BTS of a neighbouring cell. The phone and the BTSs of the neighbouring cells monitor each other others' signals and the best target candidates are selected among the neighbouring cells. This tight coupling between mobile and BTS makes handover to happen within few milliseconds hence user will not notice any change in signal while on call.

GSM systems use received signal power and signal-to-noise ratio along with other parameters to decide a handover, CDMA systems use received signal power, bit error rate (BER), block error/erasure rate (BLER) and other parameters to decide handover. CDMA systems also have connections with many cells at the same time by use of rake fingers. By using 3 or more fingers CDMA system can improve the quality of call.

CHAPTER 5

VIDEO OVER IP

5.1 Overview of Video over IP

Video over IP networks has been increasing swiftly for many years due to growth of the Internet technology and high-speed networks. Also we have many standardized IP protocols to transmit video over IP networks [50]. Video over IP (VVoIP) is the emerging paradigm of VoIP which has emerged as a successful technology as it offers low cost, flexible, efficient network utilization as it uses IP network, user mobility as well as number portability and ease of integration with cellular and PSTN networks. VVoIP emulates live human interaction, thus offers huge potential in revolutionizing the future modes of communication than VoIP. Advents in the last mile access technology have enabled users to experience higher bandwidth as well as emergence of IP based like SIP and IMS, are paving the way for realization of VVoIP based services. Though VVoIP offers similar technological challenges as VoIP for delivery of services over the network, it has much higher and critical constraints [37].

Below we discuss some of its example uses.

Example 1: Secure monitoring of different video source over the LAN/WAN/WWW with IP packets terminating on a PC

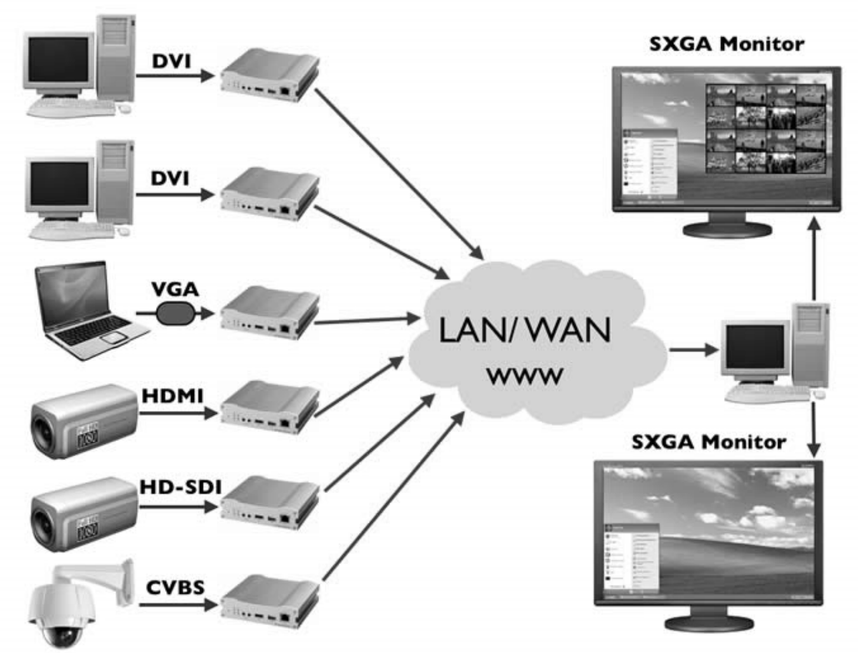

Figure 5.1 Video Over IP: Example-1

Example 2: Transfer of video streams directly to display screens over IP packet network

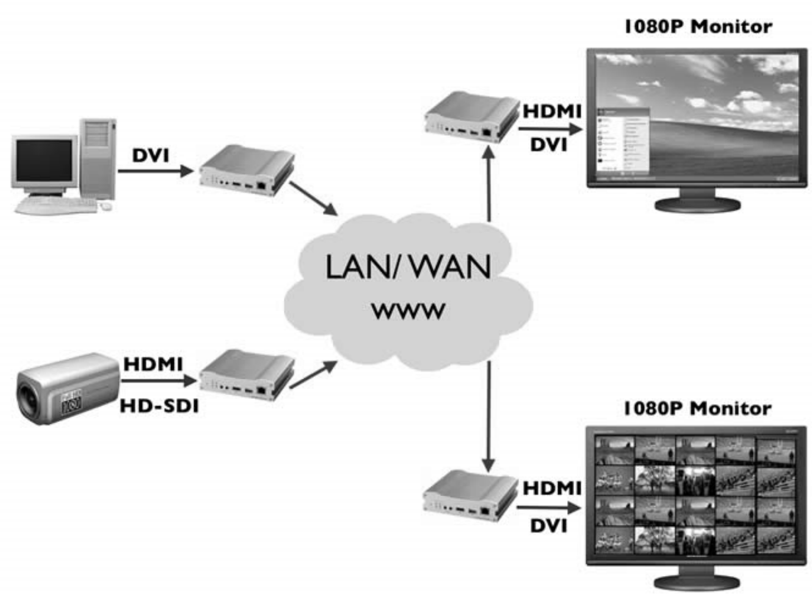

Figure 5.2 Video Over IP: Example-2

Example 3: Supporting multiple devices with different capabilities

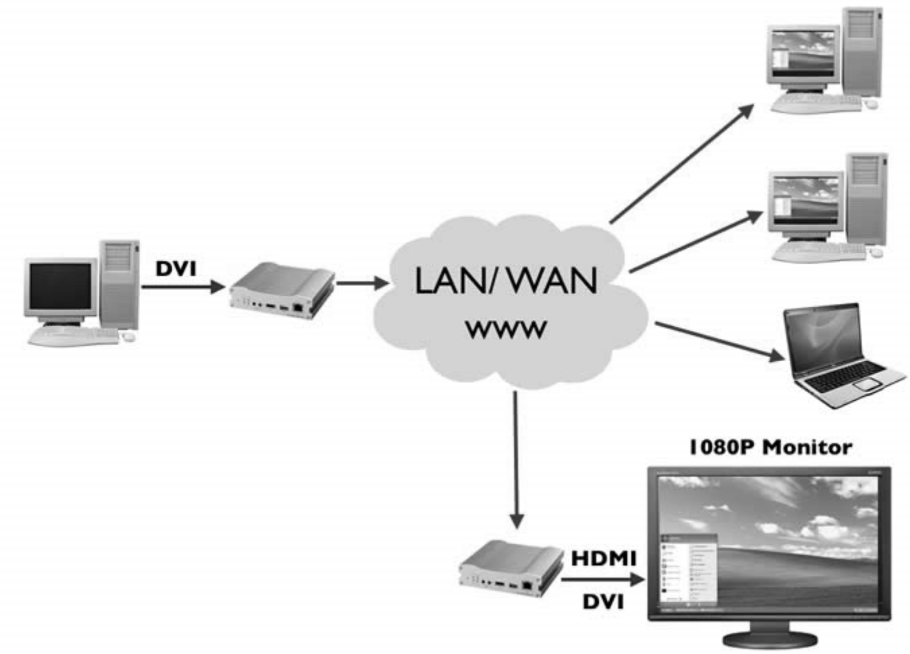

Figure 5.3 Video Over IP: Example-3

In this example both PC based decoding and an HD server configured as a decoder, with no PC involved.

Applications: Hospital Endscopy, MRI and CT scans plus other PC based graphics can be sent over the IP network. Two way audio and full HD Video opens up possibilities of consultant surgeons advising theatre surgeons remotely. Training of students showing live situations in the field whilst students are located remotely are possible [52], HD content allows them to see what the trainer sees.

5.2 Pixels, Luma, Scanning and Chroma

It is the basic building block for all forms of digital imaging, including both still photography and video. Every picture is made up of innumerous number of pixels depending upon the quality of picture.

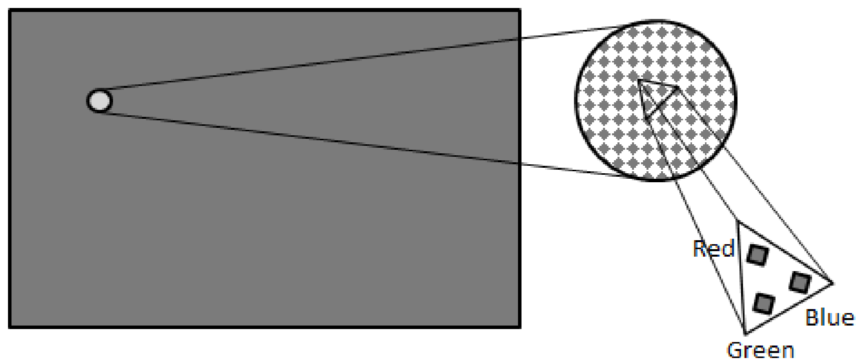

Figure 5.4 Typical Dot on Television Display

Luma represents the brightness of each pixel in a video signal.

Scanning is used to capture, store, transport, and display the luma and chroma values of each pixel. A video terminal uses this information to display images in the order

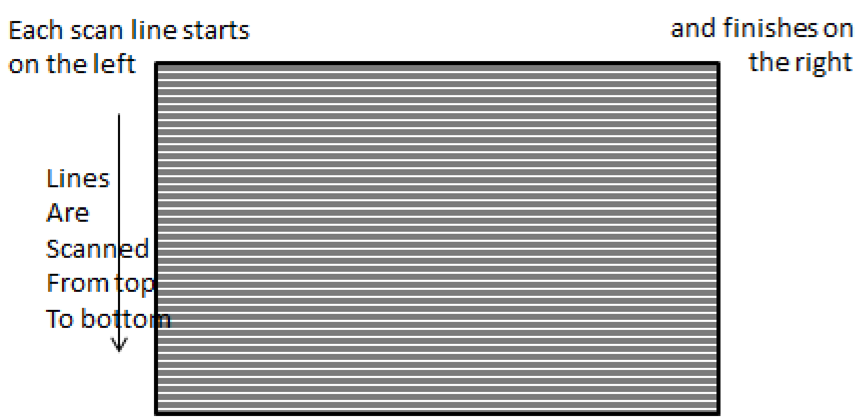

Figure 5.5 Video Display Scanning

Chroma is the portion of a video signal that represents the color of each pixel. Colors are intended to range over the full spectrum of the human visual system, from red through green and blue, in any combination.

5.3 Types of Video

The following are types of videos found in capturing, transmission and processing

5.3.1 Composite Video

This type of signal contains all of the required information to form a full-motion, full-color image on a television screen. A single coaxial cable can be used to send composite signal between devices such as a videotape player and a television set. Composite video signals are also used for normal analog television broadcasts, and in this case they also contain audio information.

5.3.2 S-Video

S-video signals are similar to composite video signals, with one crucial difference. In S-video, luma information and chroma information are carried on different wires. This is why an S-video cable has four pins: one pair for the chroma signal (I+Q for NTSC, U+V for PAL) and another pair for the luma (Y) (plus an outer shield, for those quibblers in the audience).

5.3.3 Component Analog Video

Component analog video offers benefits over composite and S-video. Because different color signals are carried on different sets of conductors, processing of the signals is kept to a minimum.

5.3.4 High-Definition Video

High-definition video (also known as high-definition television, HDTV, or simply HD) offers much more detail in video images because it uses many more pixels than standard video. These added pixels allow much larger video displays to be used without the loss of sharpness that can come from simply enlarging a standard-definition video signal.

5.3.5 Serial Digital Video

Within the world of professional video production, virtually all new equipment installed today is digital. Video cameras produce digital outputs that go to digital tape recorders or digital video. Videoediting suites have also gone all digital—video content is converted into digital form and manipulated using software.

5.3.6 Consumer Digitial Video

SDI and HD-SDI signals are not used in a typical consumer's home. Instead, Digital Visual Interface (DVI) and High-Definition Multimedia Interface (HDMI) connectors are used. A big advantage of these connections is that they are bidirectional, which allows a video source to determine the capabilities of the display device and also to encrypt the signal. Both these capabilities are crucial to the success of digital television, because scaling of 720p and 1080i signals is required to display correctly on the many different pixel counts used in consumer screens, and content security is required by the providers of HD content.

5.3.7 Internet Video

Internet video is quite different from broadcast television in the user experience. Not only do viewers have more control over the video playback, but they are also required to seek out and select the content that they want to watch. This is quite a contrast to the continuous 7 x

24 streams of broadcast TV that are delivered to millions of viewers simultaneously. Two downsides to this freedom are a lack of quality control for the technical aspects of the video and audio signals as well as huge inconsistencies in the esthetics of the video pieces; many Internet viewers have become accustomed to viewing a mixed bag of worthwhile and worthless content.

5.4 Compression

When video signals are transported over an IP network, they are most often compressed. In this context, compression means to reduce the number of bits that are required to represent the video image. Video technology users are free to choose whether or not to employ compression for their video signals. However, it is important to understand that the choice of a compression method can sometimes mean the difference between success and failure of a video networking project.

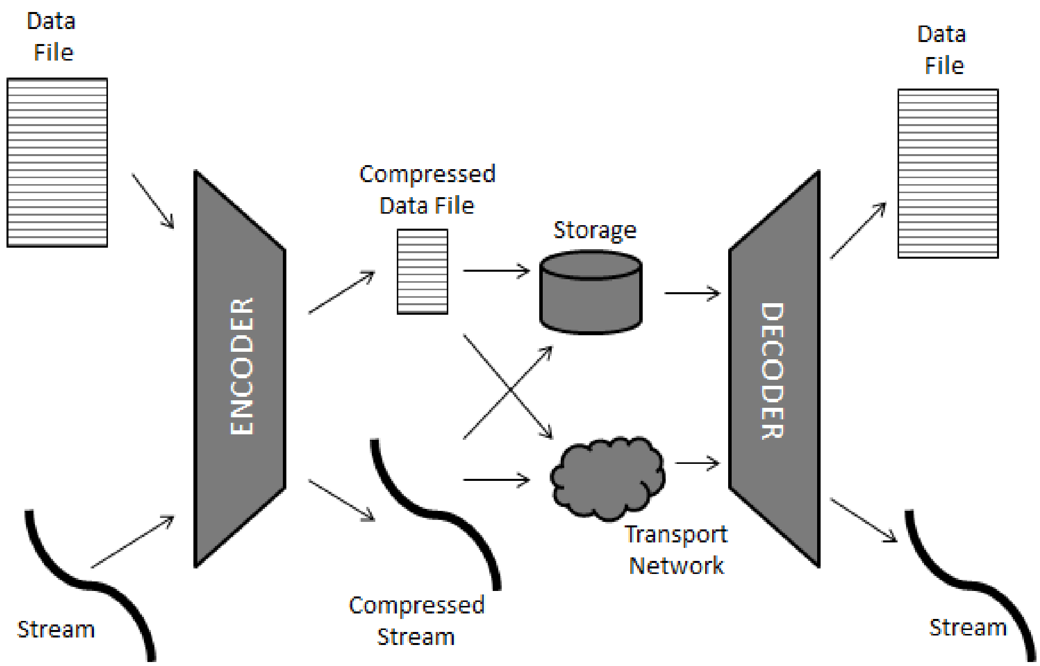

Figure 5.6 Compression Overview

5.5 Encapsulation

Encapsulation is the process of taking a data stream, formatting it into IP packets, and adding the headers and other data required to comply with a specific protocol. As we shall see, this is not simply a matter of taking a well-established formula and applying it. Rather, the process of encapsulation can be varied to meet the performance requirements of different applications and networks. Any data that is to flow over an IP network needs to be encapsulated into IP packets, whether the data is a prerecorded file or a live digital video stream. Packet encapsulation is done in real time, just before the packets are sent out over the network, because much of the data going into the packet headers (such as the destination IP address) changes for each user. Software tools to perform encapsulation are included in a wide variety of devices, including desktop PCs and file servers.

5.6 Packet Size

Performance of IP video signals will be affected by the choices that are made for the video packet size. The length of the packets must meet the minimum and maximum sizes in the specifications for IP. However, within those constraints, there are advantages and disadvantages to using long packets, just as there are advantages and drawbacks to using short packets. Figure 5.7 shows how the percentage of overhead changes with packet length. There is no simple recipe for choosing a packet length, but here are some of the advantages of choosing long packets, followed by the advantages of choosing short packets:

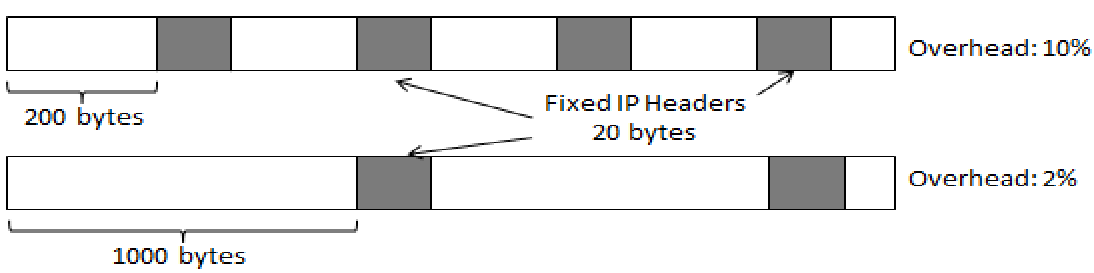

Figure 5.7 Short vs Long Overhead Comparison

Long Packet Benefits

Less overhead: Long packets and short packets both require the same number of header bytes to comply with IP. However, a longer packet will have less header information as a percentage of the overall packet size, since the packet data is longer. In other words, the total amount of network bandwidth necessary to transport a given video stream will be lower for long packets than for short packets.

Reduced packet-processing load: Each time a packet is processed by an IP client or router, the packet header must be examined, the header checksum must be verified, acknowledgments may need to be sent, and so on. Servers, routers, and client PCs must do this work for every packet, whether the packet is large or small. With larger packets, the amount of work is reduced, which can result in smoother operation of the network.

Greater network loading: In Ethernet and other network technologies, short gaps must be present between data packets sent by a transmitter. With longer packets, fewer gaps are required, meaning that more of the link’s bandwidth can be used for user data.

Short Packet Benefits

Lost packets are less harmful: In the event that a packet suffers a bit error in the header, the packet will be discarded. With short packets, each lost packet contains less of the overall video data. With some video streams (particularly ones that contain errorcorrection data), the video playback device may be able to recover from data loss or at least mask it from the viewer. As packet lengths go up, each lost packet means more lost data, which in turn makes it harder to recover from lost packets or other network impairments.

Reduced latency: When a video stream is sent out over IP, each packet cannot be sent until it is full of data. With very lowbit-rate signals (such as a web camera with a very low framerate or an audio-only signal), the amount of time it takes to fill up a long packet can add to the end-to-end delay of the transmission system. Also, some error-correction techniques process multiple packets in one batch; these packets will be accumulated faster (and latency reduced) when short packets are used. Voice-over-IP packets typically use short packets to reduce end-to-end delay.

Less need for fragmentation: Whenever a packet is too long for a particular network segment, it must be broken into smaller packets by a process known as fragmentation. This action can create significant overhead for routers and other devices along a network and should be avoided. The best policy is to use a packet size that is less than the maximum packet size allowed by any segment of the network between the video source and the video destination.

There is no single answer for the question of optimum packet length. Typically, video signals tend to use the longest possible packet sizes that won't result in fragmentation on the network, to minimize the percentage of overhead. A mistake in selecting a packet size will generally not prevent video from flowing, but it can create extra work for devices all along the path of a connection. The best thing for a user to do is to start with a packet length that is reasonable, given the network conditions, and, if performance is below expectations, to test different packet lengths as needed.

5.7 Mpeg Stream Types

Elementary streams are the raw outputs from MPEG video and audio encoders, and they are the standardized input for MPEG decoders. These streams contain only one type of content, so at least two elementary streams are required to produce an output that has both sound and an image.

Packetized elementary streams are easier-to-handle versions of elementary streams and contain timing information that allows video and audio streams to be synchronized. A packetized elementary stream can also be made up exclusively of data such as closed captioning information.

Program streams combine several types of packetized elementary streams (video, audio, and possibly data) to support production and recording tasks, and they are used on DVDs. Transport streams are another way of combining several packetized elementary streams into a single entity that can be transported across a network. Transport stream packets have a fixed length, and the streams carry clock information that is needed for real-time signals.

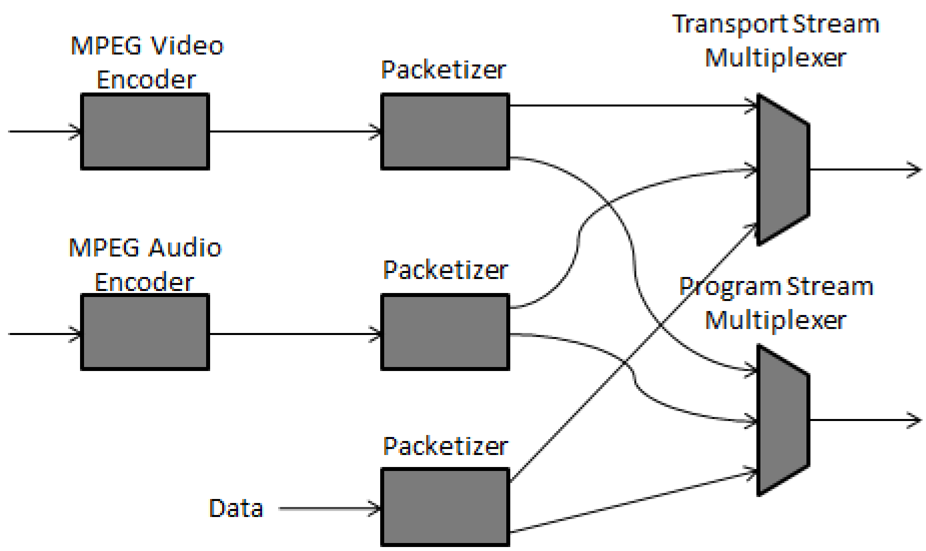

Figure 5.8 MPEG Stream Types Encoder Flow

5.8 Transport Protocol: RTP

The Real-time Transport Protocol is intended for real-time multimedia applications, such as voice and video over the Internet. RTP was specifically designed to carry signals where time is of the essence. For example, in many real-time signals, if the packet delivery rate falls below a critical threshold, it becomes impossible to form a useful output signal at the receiver. For these signals, packet loss can be tolerated better than late delivery. RTP was created for these kinds of signals, to provide a set of functions useful for real-time video and audio transport over the Internet.

5.9 Network Impairments

The following are some of the common network impairments found in a typical network.

5.9.1 Packet Loss

A packet is declared “lost” when it does not arrive at its intended destination. This can be caused by any number of circumstances, such as a network failure. It can also be caused by subtle errors, such as a single bit error inside a packet header. It can also be caused by faulty network configuration, which sends packets to the wrong destination.

5.9.2 Packet Reordering

Packet reordering occurs in a network when the packets arrive in a different order from how they were sent. For many types of data, such as e-mail and file transfer, this is not a problem, since the data is not normally used until all the packets have arrived. However, for video data streams, out-of-order packets can cause problems. Streaming video data, particularly MPEG video packets, has a very precisely defined structure and sequence.

5.9.3 Delay

Delay is going to happen in any network, whether from a desktop to a server or around the globe. There are two main sources of delay on an IP network: propagation delay and switching delay. Propagation delay is the amount of time it takes for information to travel from one location to another. Switching delay occurs at any point in the network where a signal needs to be switched or routed. Switching can occur when moving from one network to another or within a network any time a router is used to process packets.

5.9.4 Jitter

Jitter is a measurement of variation in the arrival time of data packets.

Table 5.1 Key Features of a Video Network

Feature	How the Public Internet Measures Up
High bandwidth	A video transmission requires high bandwidth, If every user assumes unlimited bandwidth of internet, then no one would be able to communicate
Low Jitter	A video transmission requires low jitter but that is unavoidable in current internet because there is no universal mechanism to make sure a low jitter in an end to end scenario
Low delay	A low delay is needed to ensure real time working of certain applications such as video conversation, robot movement monitoring, etc. but that is also unavoidable because it depends on the route taken by packets and delay in routers.
Priority control	Internet lacks any priority control
Lossless transmission	Internet does not provide 100% transmission guarantee for packet transmission. Packet losses are inherent in routes taken by packet to reach destination.

5.10 Challenges of delivering Video over IP

Delivering Video over IP offers similar challenges as Voice over IP, though the former has more stringent constraints. Below we discuss some of the challenges associated with FMC in general as well as related to delivering VVoIP.

-Access network capacity plays an important role in delivering multimedia services [51]. It is important to have an accurate analytical model for capacity estimation under diverse traffic conditions for admission control and resource provisioning. Further, it can assist in accurate network planning in the fixed network for maximum utilization of the available infrastructure, load balancing, as well as mitigate additional deployment of new equipments in backhaul networks.

-The perceived QoS requirement must be accurately mapped to network resource to enable end to end QoS provisioning. While this looks simple, we need to develop an analytical framework to accurately quantify QoS to map it to the underlying network resources. Additionally, interoperability with different network and link layer QoS mechanisms, like MPLS, diffserv, 802.1q/p, needs to be understood for efficient deployment.

-Some enterprise network having critical data application do not like to put voice and video data in the same network as data.We need to explore mechanism, for example QoS enabled overlay networks or network virtualization, to support VVoIP services over them.

-Effective quality measurement techniques have to be designed to observe traffic in real time and appropriate measures have to be taken to enhance the user's quality of experience (QoE).

-Video traffic has diverse characteristics depending on the video content and codec used. Hence cross layer information and optimization must be enabled to exploit these characteristics in configuring the network and link layer parameters for more efficient resource management and QoS provisioning.

-Based on our experiences with video codecs like MPEG-1, 2 or 4 and H.264, we need to extend them to design and enhance session layer protocols, such as H.323 and SIP, to support VVoIP as well as design mechanism for session mobility. We must also study the use of flow aware routing as well as exploit the characteristics of video traffic from codecs to efficiently create and manage flows, e.g. split flow mechanism etc.

-Instead of the end nodes being aware of the connection, the onus should now shift to the network, where participating network components are aware of the connection and assist in recovering from network idiosyncrasies (delay, jitter, losses, inter stream latency and out of order delivery) as well as QoS provisioning and flow management.

-Mobile devices are designed to be energy efficient. Delivery of high bandwidth multimedia traffic over WLAN puts extra load on the battery. Energy efficient communication mechanism for delivery of video traffic needs to be designed over different networks

-Communication and content delivery over video has primarily been passive. Clients, network and services should enable active user participation to create and control dynamic video content as well as deliver them.

-It has been well established in the literature that voice and video traffic are not friendly to some of the security measures undertaken, like firewall, NAT, etc. The session layer protocol, H.323 or SIP, used to deliver these services use features, like dynamic port based signaling orexposing private address space. Thus security mechanism need to be built on the available technologies to make them H.323 and SIP aware, thus voice and video traffic aware and help in mitigating network idiosyncrasies for efficient delivery of services.

CHAPTER 6

SOLUTION AND APPROACH

6.1 Overview of multimedia conferencing and comparisons

A multimedia conferencing architecture should allow an user to initiate a conference, to request to join an ongoing conference, to invite another user to enter an ongoing conference, to drop from a conference, to remove a participant from a conference, and to obtain data transmission privileges [53] A multimedia conferencing architecture should provide a high degree of flexibility and adaptability, the security mechanism, the integration with the existing management system, and a high scalability[54]. Most importantly multimedia conferencing architecture should support mobility. The process of mixing and encoding of different types of media on a centralized server has been studied for multimedia conferencing [55]. We propose a highly scalable and mobility enabled distributed architecture named “Call Control Network Architecture (CCNA)” based on key architectural evolution techniques proposed for next generation networks in telecommunications such as SIP, IP, MPLS, FMC, VoIP and key architectural evolution techniques proposed for next generation Internet such as TRIAD, SFS, DONA, HIP. Apart from loosely coupled conferencing where there is no role played by deployed infrastructure. The Tightly coupled architecture has a centralized conferencing server. The general conferencing architecture consists of a conference server and means for participants to be able to do join/add/leave functions. A focus is a SIP user agent which is responsible for the management of the conference using SIP signaling protocols. All requests from participants are directed to Focus, which in-turn refers to the conferencing polices stored in the media policy databases. Media policies guide the focus for the management of various conference requests from the participants. A mixer is responsible for handling the multimedia streams, and generating output streams which can be distributed to participants. A mixer can be located either in the focus or in the participant's user agent. In both cases, a mixer is controlled by the focus.

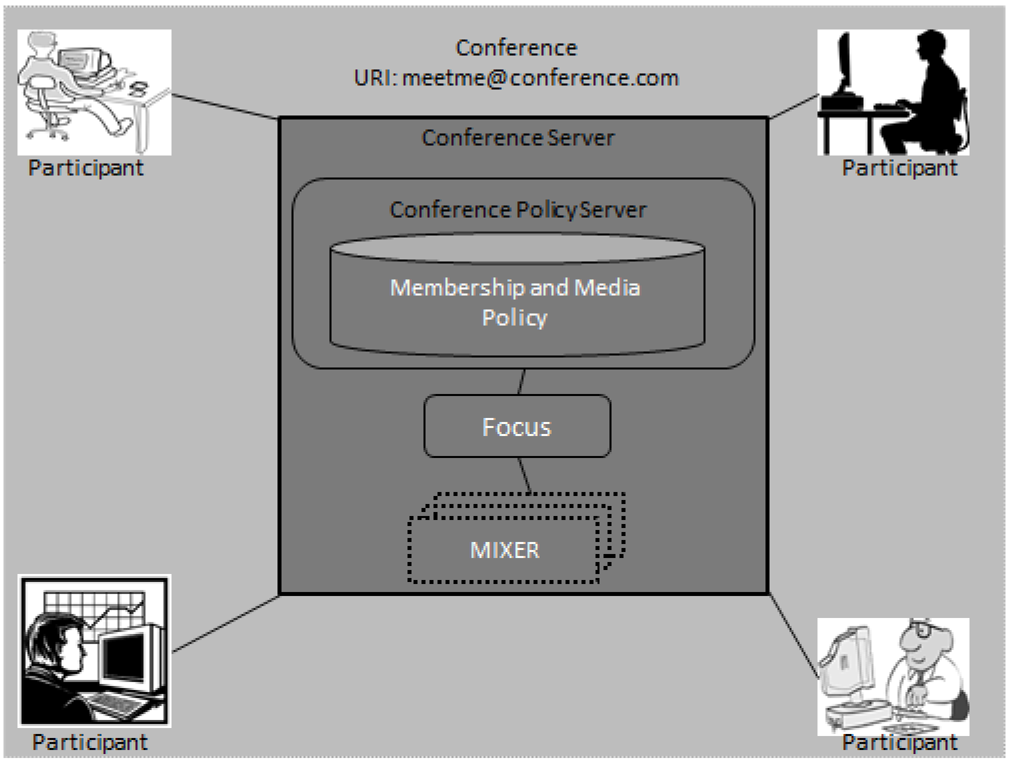

Figure 6.1 Tightly coupled conferencing

The following table compares some of existing architectures for multimedia conferencing as proposed by IEFT and others with our proposed architecture (CCNA)

Table 6.1 Comparison of conferencing models (Part A)

	Location of Focus	Location of Mixer	Number of Servers	Scalable	Built-in security
Centralized Model[62][18]	Central server	Central server	1	Medium	No
Endpoint Model[18]	Participant	Participant	0	Small	No

Table 6.1 – Continued

Media Server Component Model[18][65]	Central server	Central server	2	Medium	No
Distributed Mixing Model[18][63][64]	Central server	Every Participant	1	Medium	No
Cascade Mixers Model[18]	Central server	Distributed	Many	Large	No
Distributed conferencing Server model[56]	Distributed	Distributed	Many	Large	No
CCNA (Proposed Model)	Distributed	Distributed	Many	Very Large	Yes

Table 6.2 Comparison of conferencing models (Part B)

	Primary Focus exists	Support for middle boxes	Distribut ed policy servers	Dynamic Load Balance	Mobil ity
Centralized Model [62][18]	Yes	No	No	No	No
Endpoint Model[18]	Yes	No	No	No	No
Media Server Component Model [18][65]	Yes	No	No	No	No
Distributed Mixing Model [18][63][64]	Yes	No	No	No	No
Cascade Mixers Model[18]	Yes	No	No	No	No
Distributed conferencing Server model[56]	Yes	No	No	No	No
CCNA (Proposed Model)	No	Yes	Yes	Yes	Yes

A centralized server model has both the focus and the mixer is located in the same server. In an endpoint server model, both the focus and the mixer are located together in one of participants wherein that participant will act as both the server and the participant. In a media server component model, the focus and the mixer are located into two different centralized conferencing servers. In a distributed mixer model, which only distributes the mixer functionality among the participants but the focus is still located in a centralized server. In a cascade mixer model, the focus is located in a centralized conferencing server, and the mixer is located in several distributed conferencing servers. In another model proposed by Yeong-Hun et al[56], though focus is located on different servers based on regional focus, where each regional focus handles management of local conferencing service. There still exists a primary focus that controls the whole conference. We propose a distributed and highly scalable architecture which does not have a primary focus which controls the entire conferencing by using single policy but have multiple distributed and collaborative focus servers with separate policy servers. We believe this mode of solution respects plethora of middle boxes, policy servers, firewalls found on the Internet.

Mobility support at the core of cellular architectures is mainly supported by MSC along with Home Location Registers (HLR), Visitor Location Registers (VLR) using information stored in SIM card such as IMSI number, location area identity, secret cryptographic key etc. but in CCNA architecture mobility is supported by using dynamic registration state in network elements setup by ever mobile/roaming end points/devices.

6.2 Call Control Network Architecture (CCNA)

We propose a novel architecture having a number of connected network elements called Call Control Entity (CCE), each CCE has links to many peer CCEs and a parent CCE, as shown in below figure CCNA’s overall structure looks similar to AS hierarchy of the current Internet. Each CCE works in a similar way a SIP proxy works by routing requests, handling call setup, registration, implementing policies, authorization and network access control along with new functionalities added by our proposal such as dynamic registration state management, name based routing, control plane and media plane support for different conferencing modes, and call load distribution.

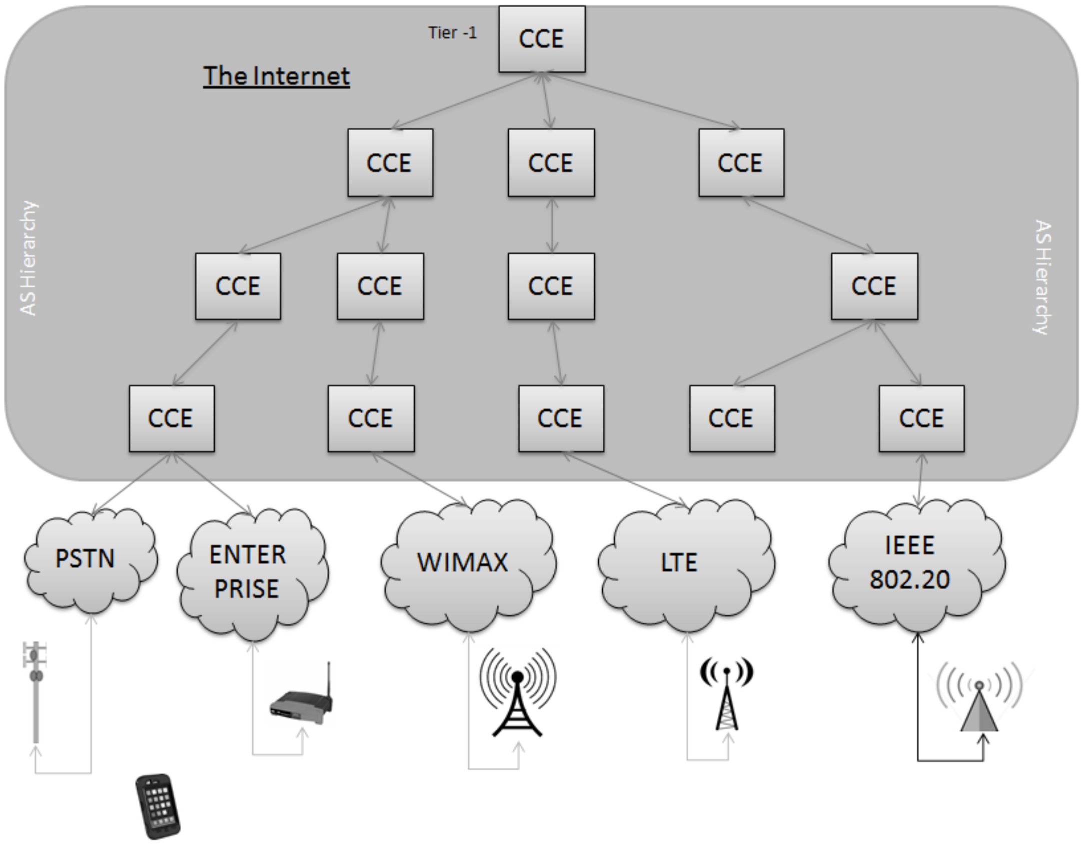

Figure 6.2 CCNA Architecture

Mobility Management techniques like vertical handoff and fixed-mobile-convergence are implemented in CCEs which are deployed at edge of internet along with support from mobile endpoints to provide seamless roaming across heterogeneous networks.

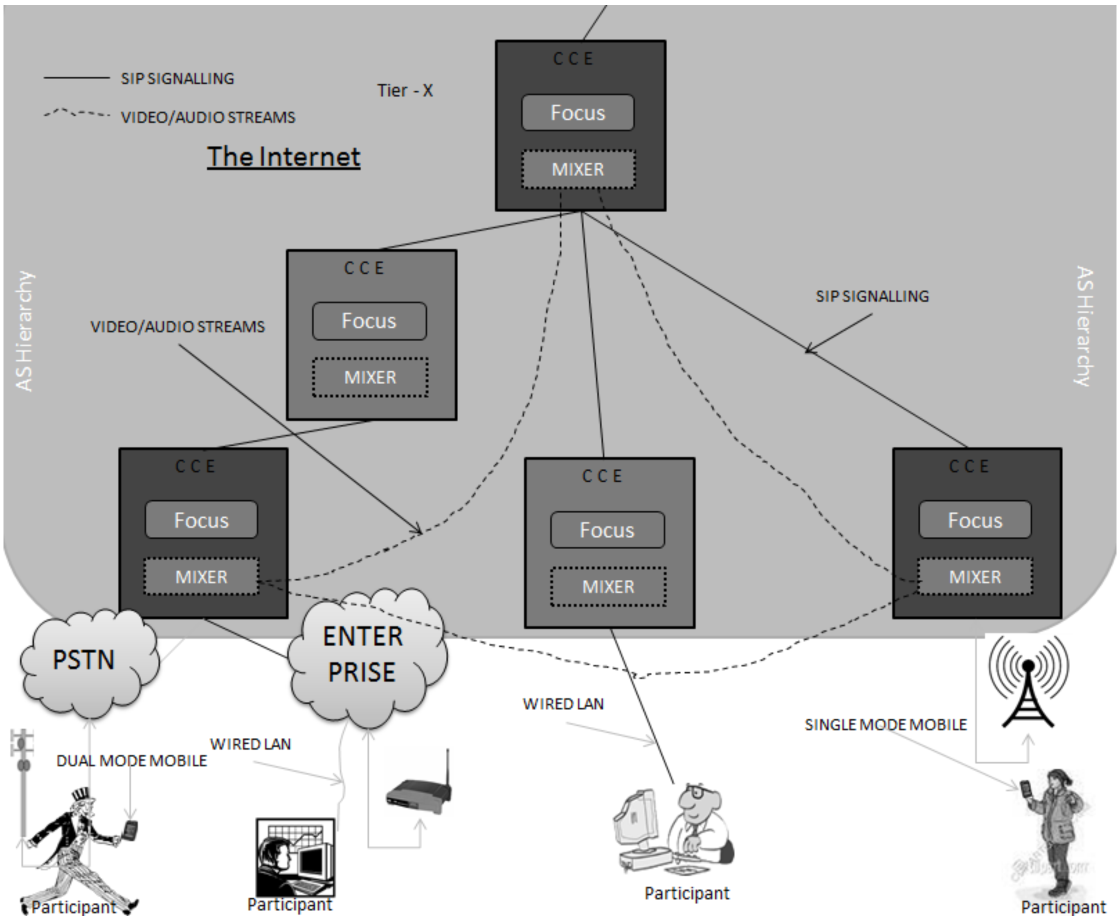

Figure 6.3 Distributed conferencing in CCNA

We will briefly describe some terms used in architecture.

Host CCE – This is a CCE which facilitates communication between a mobile endpoint and core of network. Any initial signaling messages must pass through it, a mobile endpoint broadcasts periodic SIP REGISTER message to register its location, capabilities, and other information. The CCE which is closer will be acting as Host CCE. If path based switching is used then any audio/video stream must pass through Host CCE to reach destination. Since a mobile endpoint roams it changes its host CCE also changes. A mobile endpoint can have only one Host CCE.

Anchor CCE – This is a CCE which acts as anchor point for a few (or all) of participants in the conference. It provides media plane and control plane functionalities by performing transcoding/ transrating/ media mixing/ conference control/ switching and others. A mobile endpoint can be hooked to only one endpoint for a conference. A conference can have multiple Anchor CCEs

Target CCE – This is a CCE which is selected by a load distribution and target selection algorithm to act as Anchor CCE. A Target CCE is selected based on current load and traffic conditions. A Target assumes the role of Anchor CCE immediately after its selection, but it can delegate its responsibilities to other CCE by announcing.

Mobile End-Point – This is a device user carries which communicates with CCEs through access networks such as PSTN, Wi-Fi, LTE, WiMAX. Mobile endpoints are equipped with front facing cameras and display. Whenever we specify mobile endpoint it can also mean a stationary endpoint such as PC connected over wired LAN, equipment with audio/video capture and transmission, camera, seminar room boards.

Inter-CCE Mobility – This is when a mobile endpoint roams between CCEs, this can happen when a mobile endpoint changes logs on to a access network which is under a different CCE. This mobility is provided by use of dynamic registration state maintained at the core of network.

Intra-CCE Mobility – This is when a mobile endpoint roams within access networks connected to the same CCE. This mobility is provided by Fixed Mobile Convergence and Vertical Handoff Techniques.

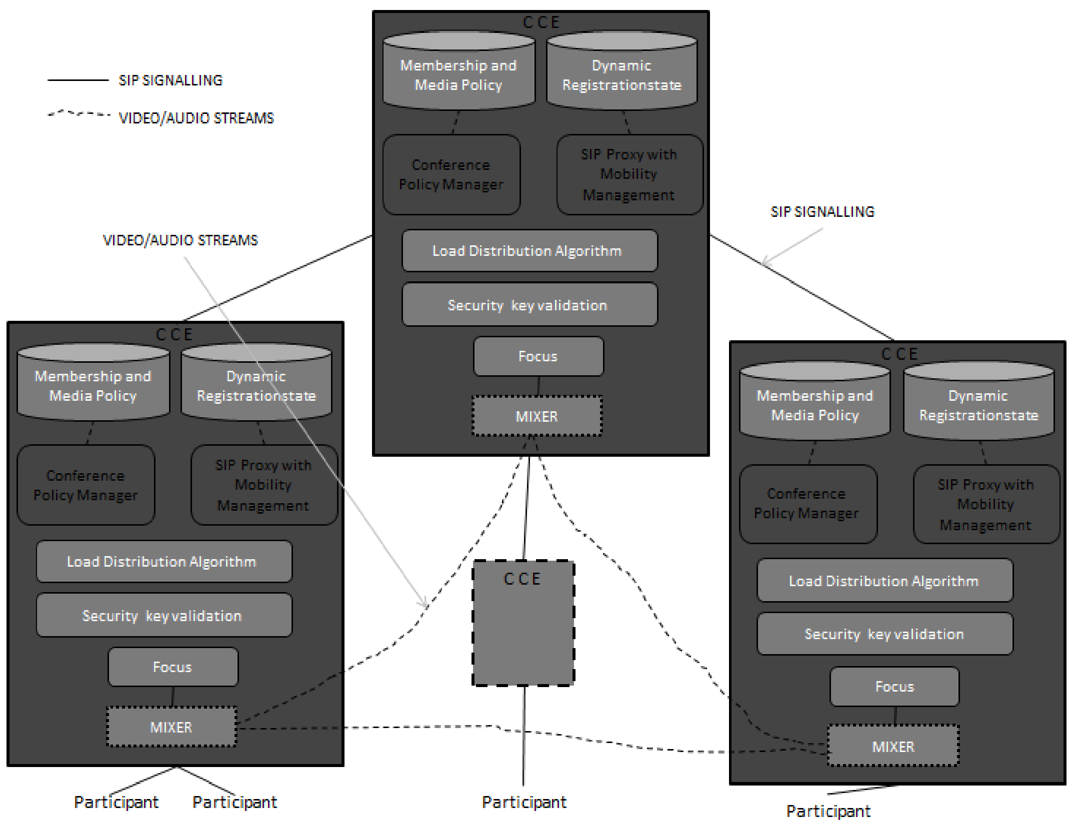

Figure 6.4 Closer look at the involved CCEs in the conference

6.3 Mobile-ID and Registration state

A mobile-ID is an identifier used to identify a mobile end point; it provides architecture a basis to construct routing, communication and location algorithms. Establishing registration state is based on mobile-ID associated with each mobile endpoint. Below we will discuss more in detail.

6.3.1 Mobile endpoint identification (Mobile-ID)

Rather than identifying a mobile endpoint using a permanent home address and a care of address as done in Mobile IP, CCNA identifies each mobile endpoint using flat self-certifying names. This avoids triangular routing problem, tunneling/de-tunneling overhead associated with

Mobile IP. Every user is associated with a public-private key pair and names are of the form "P:T", where "P" is the cryptographic hash of users public key and "T" is the unique tag associate with each mobile device owned.

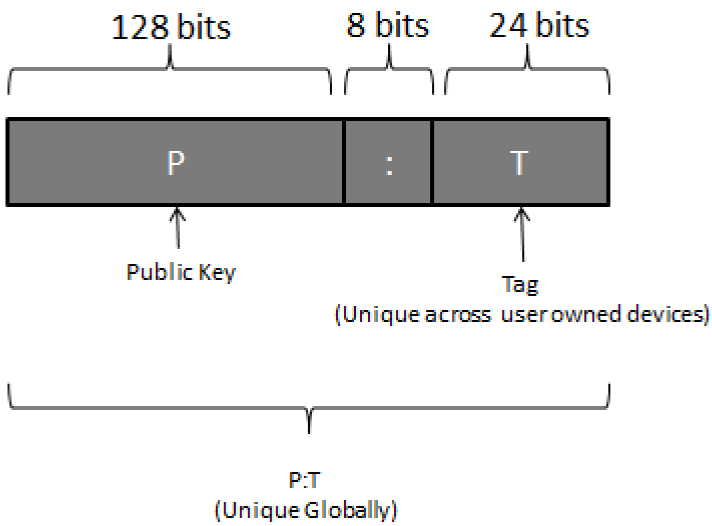

Figure 6.5 Mobile-ID

These self-certifying names are application independent and globally unique. So when a mobile registers itself with CCNA, it can be checked for authenticity by generating public hash key from P. Mobile device registers with any location using same mobile-ID/self-certifying name which enables CCNA to provide mobility based on dynamic registration state set up by roaming/mobile endpoints.

This flat-name raises the question about how user remembers it to perform operations, for example - to call someone. We believe that mapping to *mobile-ID* needs to be done using some external mechanisms familiar to users domain. Below are some examples that user want to map

Mobile-ID = zc45dopsd5jksep:ab3

"+01 682-553-1142" -> "zc45dopsd5jksep:ab3" (in this case we need mechanisms to map numbers)

"Jimmy" -> "zc45dopsd5jksep:ab3" (in this case we need mechanisms to map user friendly names)

6.3.2 Dynamic registration state setup

SIP REGISTER messages are used to setup the registration state. A mobile endpoint broadcasts register messages periodically to update its location, status and other information to its nearby CCE. which in-turn will process REGISTER message using following algorithm as shown in Figure xx which includes determining its authenticity, if any update required, if needs to be forwarded this message to its parent & peer. Each "CCE" maintains a registration table that maps a mobile-ID to its next-hop CCE, distance to mobile endpoint (in terms of the number of CCE hops or some other metric), capabilities and other information as shown in table 6.6.

Mobile-ID	Next-Hop	Key	Capability	Other info
xxxxxx:xx	IP of next HOP CCE	<security key>	<video/audio/codecs>	
xyxxyy:xy	129.107.908.444	xxxttys	G.711, H.320	
xyxxyy:xz	129.107.908.444	xxxttym	G.729, H.323	

Figure 6.6 Registration table

Below we show the register message which is a modified SIP REGISTER message containing mobile-ID and other CCNA related information

REGISTER 43c52e9d29317c0b:ab3
Content-Length: 0
Call-ID: ED9A8038-A29D-40AB@ab3
CSeq: 36 REGISTER
From: <43c52e9d29317c0b:ab3>
Max-Forwards: 0
To: <cce_12b456>
User-Agent: SipDroid/1.5.2
<status, capabilities and other information>

Below is the Registration algorithm which handles incoming register message

```
if REGISTER received then
        if (duplicate or invalid signature)
              return;
        end if

        if (time in message is new or route cost is less")
             update mobile-ID
             set TTL for mobile-ID
             set otherinfo for mobile-ID

         end
        if (REGISTER message is new)
              add mobile-ID to table
              set TTL for name (P:I)
       end
        foreach (parent link and peer links)
          if (send-decision = true for a  link)
              create-new-message for that link
              add link cost
              sign message(private key, msg);
              send message
          end if
        end for

end if
```

A REGISTER message is forwarded based on local policy, if a REGISTER is received from child it is forwarded only when a entry in table does not exist for that mobile-ID or if the cost in that route is less. if it receives from a peer, it checks whether it can serve as a transit for signaling path/data path. Since REGISTER messages are forwarded based on local policy, CCNA takes into account the interdomain policies as specified in BGP. Even the messages are checked for authenticity when a CCE receives it by issuing a challenge with nonce. since the REGISTER messages travel along the policy driven path, middle boxes can be supported which implement various policies for a enterprise/network. a REGISTER message also specifies the status of mobile end point, whether if it is in a call or not, if endpoint specifies that it is in a call then host CCE will send INVITE to anchor CCE as specified in the REGISTER message to establish the call leg with the Anchor CCE, once this Anchor CCE receives INVITE message it will run the load distribution algorithm to determine the target CCE which is optimal based on current load and reachability. if a new Anchor CCE is selected, then mobile end point is notified of the change. REGISTER message has a TTL; this is required to maintain the current location of mobile. Even though mobile endpoint broadcasts periodic CCE messages, the receiving CCE will forward it only when there is a change in cost involved reaching it or if it is a new mobile endpoint under its locality or based on local policy

6.4 System Design of CCE

A CCE Network element can be a single/multiple server or a Cloud. Some of its functions include SIP Proxy, SIP Registrar, SIP conference server, Multipoint control unit (MCU). and ISDN Gateway (if connected to a PSTN network). The following figure shows some modules based on different operations performed. The conference control module processes the signaling messages and designed on SIP proxy, registration, load distribution message, security key validation and media plane modules. It admits incoming request to join a conference, creates and processes requests to and from other internal components. The FMC techniques can be present in any CCEs, which will enable it to handle mobile endpoints which are not capable of handling SIP.

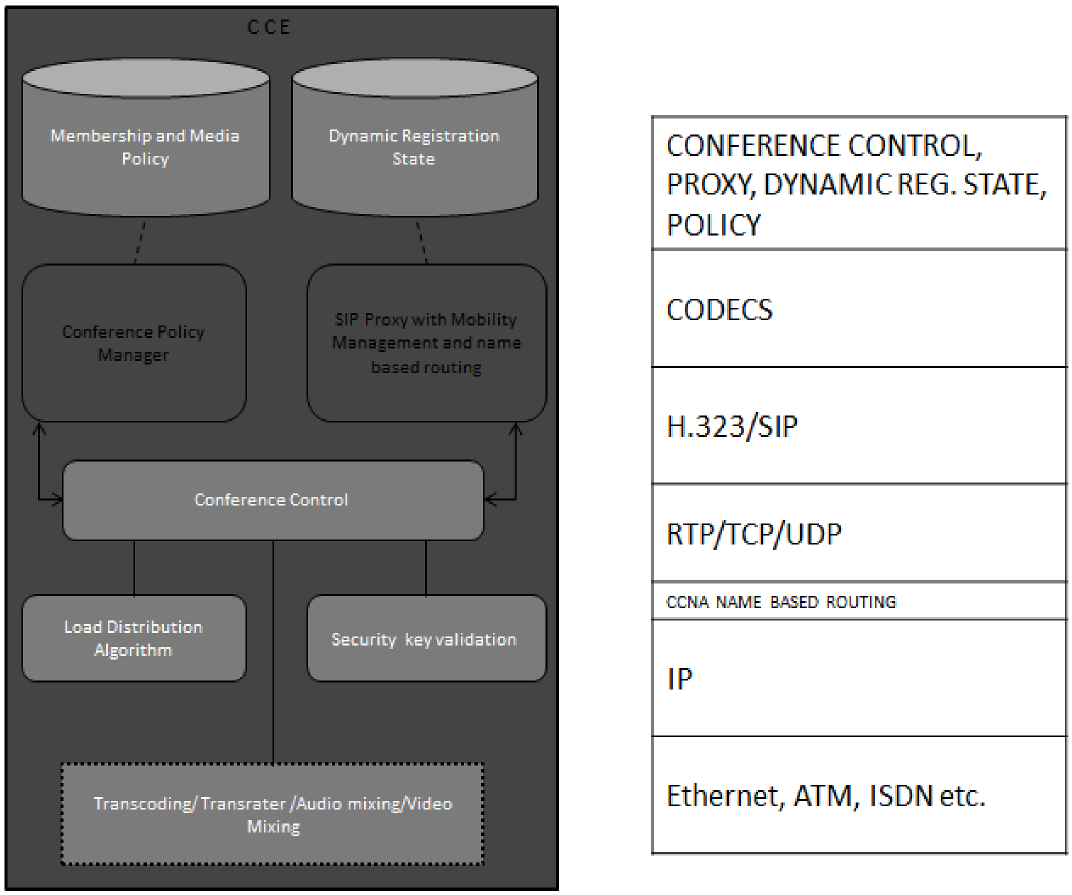

Figure 6.7 CCE Network element design

The medial plane module interface provides an application programming interface (API) for the conference control module to access the media services such as video/audio mixing. The media components send and receive audio and video packets and provide media services such as audio or video mixing. The media components can be either software modules or DSP cards co-located in that server. Because mobile endpoints may have video streams with different stream characteristics from other endpoints (codecs, bit rate, frame rate, picture size), the CCE might need to convert the video streams, depending on the endpoints’ specific receive capabilities.

Few modules based on operations performed are

- Conference Control module
- SIP Proxy module
- Registration state module
- Load distribution and target selection module
- Media plane control module
- Policy module
- Security key module

Feasibility of maintaining registration state at each CCE

Although the conferencing itself occurs in distributed way by involving multiple anchor CCEs, a CCE would need to maintain a registration state table consisting of one single entry for each mobile endpoint below or equal to it in AS hierarchy. Thus the toughest requirement would be on CCE at Tier-1, if a single entry of data for mobile endpoint is 100 bytes, it would account to less than 2-3 TB of data even if all the people of this planet (6.8 Billion) are active at the same time. The existing state of the art routers, servers and data centers can easily handle this load. A Tier-1 CCE is also void of any media plane handling functionalities which gets delegated to CCEs at lower level and having less load when CCNA runs load distribution algorithm.

Fault Tolerance and Load distribution

CCNA architecture is more resilient to fault tolerance compared to other models by design, in the case of failure of anchor CCE only the mobile endpoints anchored to it will be unable to communicate for few milliseconds then these mobile endpoints issue SIP INIVITE message which will be routed via new route and conference gets established again. Load distribution and target selection algorithms ensure that at no point a CCE is selected that has crossed its threshold operating load. A CCE is also equipped with delegation mechanism which will enroll a new CCE to take over as anchor point.

6.5 Supporting conferencing operations

As with any multimedia conferencing architecture CCNA should support user to initiate a conference, to request to join an ongoing conference, to invite another user to enter an ongoing conference, to drop from a conference, to remove a participant from a conference, and to obtain data transmission privileges and other operations below we explain some of them in detail.

6.5.1 Basic Conference Initiation

Once the devices setup its state with CCEs, a basic two way video session starts by a device user sending SIP INVITE(P:I) and routed by CCEs to its current location. Once INVITE(P:T) reaches destination device it starts negotiation with originating device for capabilities and starts sending/receive of video data over IP. If a mobile endpoint is not capable of SIP then local CCE which is near to it acts as ISDN gateway to bridge different technologies and convert messages between SIP and other formats.

INVITE zc45dopsd5jksep:ab3
From: <43c52e9d29317c0b:ab3>
Call-ID: ED9A8038-A29D-40AB@ab3
CSeq: 1 INVITE
Content-Type: application/sdp
Content-Length: 611
Media Description,name and address (m): audio 5000 RTP/AVP 3.0
Media Attribute (a): rtpmap:3 GSM/8000
Media Attribute (a): rtpmap:0 PCMU/8000
<conference mode information>

We will discuss more about supporting devices not capable of SIP in section 6.7.2. When a INVITE comes the forwarding rule is straightforward if there is an entry in the registration table, the INVITE is sent to the next-hop CCE (and if there is more than one, the choice is based on the local policy and which entry is closest); otherwise, the CCE forwards the INVITE towards its parent (i.e., its provider) using its local policy. Thus, registration table misses are forwarded up the AS hierarchy in the hope of finding an entry.

The INVITE message does not just resolve the name, it initiates the transport exchange. The name-based routing provided by CCNA ensures that the packet reaches an appropriate destination. If the INVITE request reaches a Tier-1 AS and doesn't find a record associated with that mobile-ID, then the Tier-1 RH returns an error message to the source of the INVITE. If INVITE reaches a mobile endpoint then CCE servers involved in path decides the target CCE to handle audio/video mixing and switching functionality by using a load distribution algorithm.

When a conference is initiated by a mobile endpoint the load distribution algorithm is performed by host CCE. If a mobile endpoint joins an ongoing conference then load distribution algorithm is run by its anchor CCE.

```
/*processing OK message*/
if OK received then
        if( host CCE or nearby anchor CCE)
                if (mobile endpoint status == not in conference)
                target CCE = load distribution algorithm
                append OK with target/anchor CCE info
        end if

      /*forwarding OK*/
      append OK with current load information, link cost and hierarchy level number
      /*if current CCE == host CCE then fowarded OK reaches mobile endpoint*/
   forward OK to next hop CCE
end if
```

```
/*processing ACK message*/

if ACK received then

    if (anchor CCE is current CCE)

    invoke media plane to handle tasks

    /*Forwarding ACK*/

    forward ACK to next hop CCE

end if


/*load distribution algorithm*/

 for each pair of endpoints

        for (x=0; x <number of CCEs;x++)
          cost = CCE_load + Intracost
          if (old_CCE hierarchy > current_CCE hierarchy)
            add hierachy load
         end for

         good_cost=cost/2;

        select CCE based on good_cost from x=0
 end for

 select the CCE with maximum occurrence between all the pairs
```

Target CCE selection is based on runtime overloading data exchanged between CCE's, policy of CCE and transcoding/transrating capability request by involved mobile endpoints. This selected target acts as focus, mixer and anchor point for the current conference. Now mobile end points can send the audio/video streams to selected Target CCE in any path and need not go through all the involved CCEs.

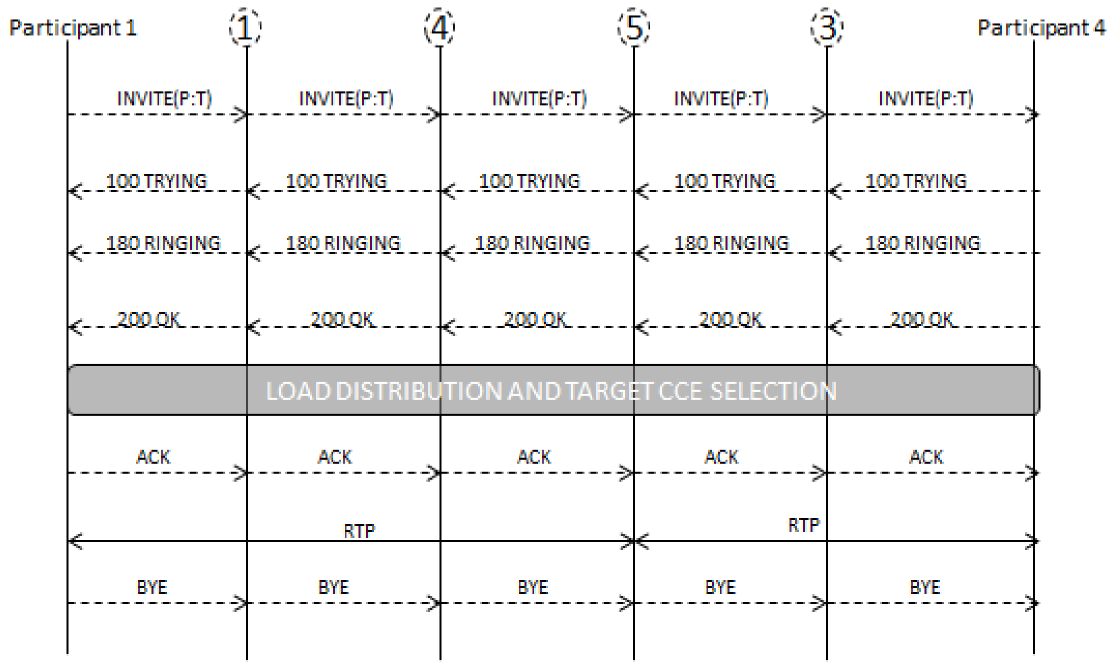

Figure 6.8 Two-way conference setup

6.5.2 Upgrading to multi-way video conference/Inviting users to join

If any mobile endpoint involved in a two-way wants to invite a new mobile endpoint, he will send a new SIP INVITE message to the target CCE, which in-turn will forward INVITE message based on registration state which will be routed to that new mobile endpoint. Once the INVITE message reaches new mobile endpoint, it will send OK message to the target CCE, which will run load distribution algorithm to determine if another target CCE should be elected to perform mixing and be involved as another anchor point or it wants to delegate its current responsibilities to newly elected target CCE resulting in a single anchor point for 3 participants in which case needs to inform all three participants about new target CCE. If newly elected target CCE only involves as a distributed CCE then only newly joined user needs to be informed about where to send/receive streams. This procedure repeats if another user device needs to join the conference.

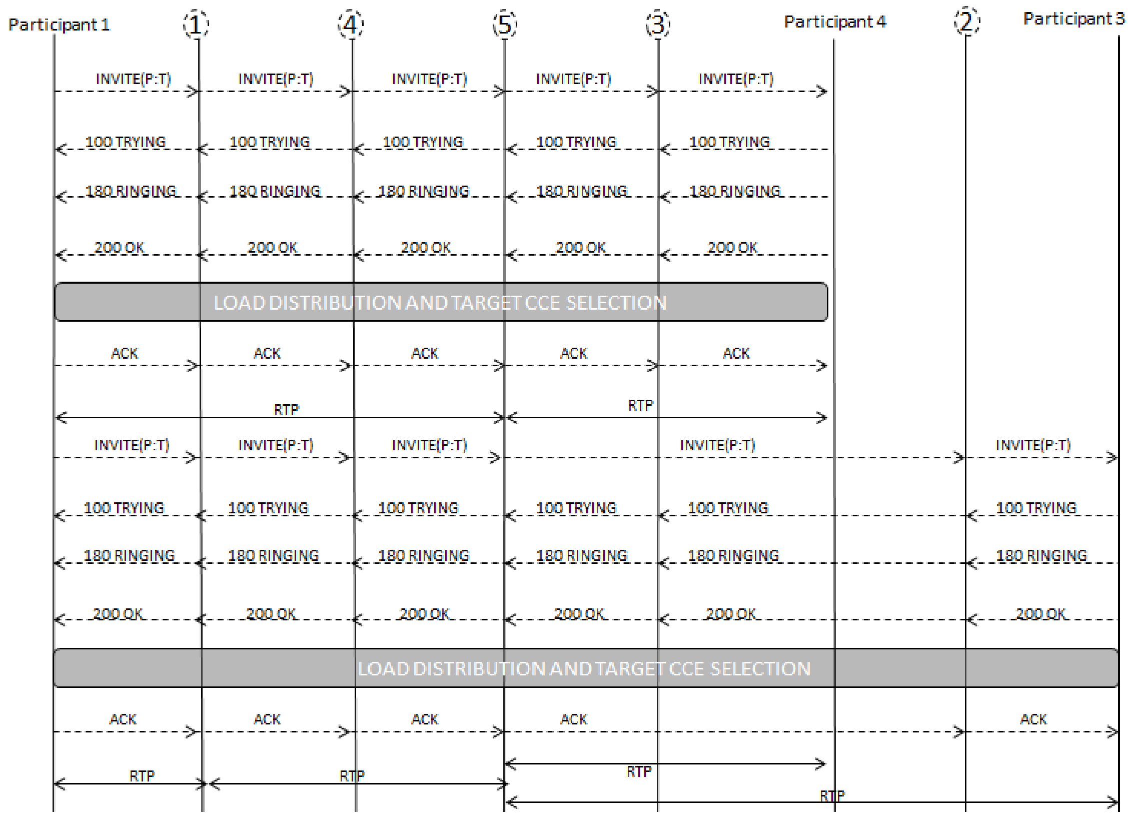

Figure 6.9 Upgrading to 3-way conference

6.5.3 Conference termination/User exit

If any mobile endpoint wants to terminate the conference, all it has to do is send SIP BYE message to its anchor point. After a SIP BYE is received at CCE it will check if conference session needs to be terminated based on the number of mobile endpoints in the conference at that moment. If the number of mobile endpoints anchored is less than or equal to 1 and current CCE is not involved in session with another target CCE it will end the session. If the number of mobile endpoints anchored is equal to 0 after SIP BYE received and processed, it will check if its involved with any target CCE if not it will end conference session, if involved then will send the involved CCE that it need not send/receive streams to it.

6.6 Supported video conferencing modes

CCNA can support all popular modes by collaboration of involved CCEs in the conference. The mobile endpoint can also specify the mode in the initial SIP INVITE message.

1. MCU Switched and Voice Activated conferences
2. Continuous presence conferences
3. Lecture mode and Round-robin conferences

6.6.1 MCU Switched or Voice-activated conferences

The different types of conferences in this mode are

1. Chair or lectern control: A single endpoint is designated as the chair of the conferences, and it controls what the other endpoints can see. This is particularly useful in distance learning environments. The mobile endpoint which initiates the conference specifies in SIP INVITE message about mode of operation the first anchor CCE selected along with other anchor CCEs involved will ensure this mode.

2. Follow the speaker: The CCNA attempts to send video from whichever endpoint is currently speaking to all of the other conference participants. This is by far the most popular conferencing mode used and we will explain how it is supported in the following section.

3. User control: Each endpoint can select any one of the other locations to watch. This places a burden on the end user and can be hard to keep up during a long conference. Some users consider this as annoying as watching someone change channels repeatedly on a television set. Mobile endpoint communicates with its anchor CCE to switch locations and in this mode each CCE receives all the streams all users but sends only one stream which the mobile endpoint requests.

4. User-controlled broadcast: A user can request to send his or her endpoint's video to all of the other endpoints in the conference. This might be used, for example, when one party is giving a presentation.

Explaining "follow the speaker" mode

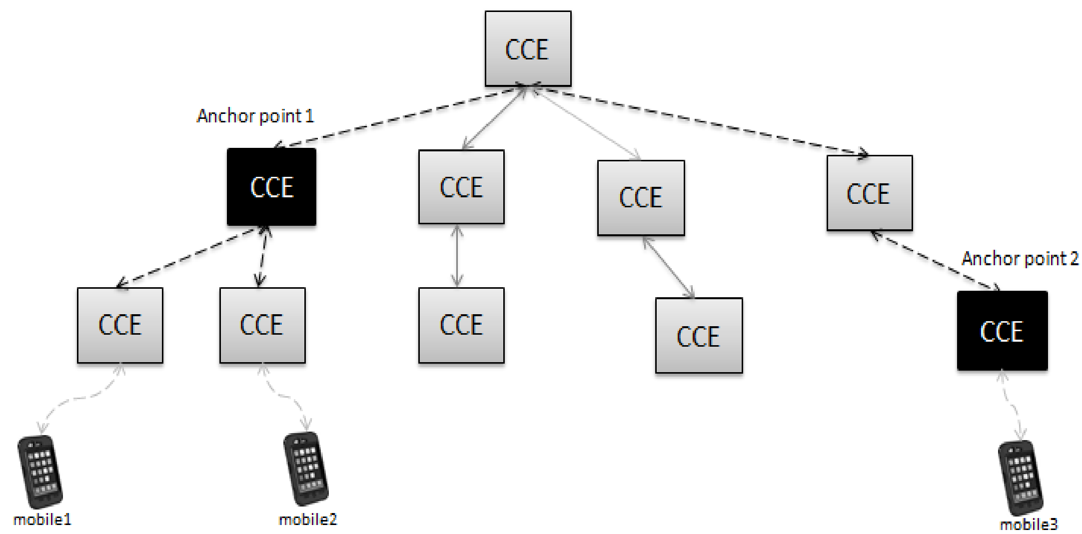

Figure 6.10 "Follow the speaker" example

The above figure shows that 3 participants are in the conference with 2 anchor CCEs. In the above figure the flow of video and audio streams are shown by dotted lines. The path of flow of streams can change depending on network topology and traffic conditions between mobile endpoints and anchor CCEs. To begin with when mobile-1 starts speaking and transmitting streams to Anchor-1 (Anchor point 1) compares the voice level of incoming streams from both mobile-1 and anchor point-2, it determines to go ahead and transmit streams to anchor point 2, once anchor point-2 receives the streams it will compare voice level and forwards streams to mobile-1 and mobile-3. But after a while if mobile-3 starts speaking and sending streams. Now anchor point-2 determines to stop receiving streams from anchor point-1 and starts sending streams to anchor point-1 and once anchor point -1 receives streams it will compare voice level and starts to transmit streams to mobile-1 and mobile-2. The advantage of this is in case of centralized server which is overloaded with all the streams, but here streams are handled by multiple servers which inter communicate to support the conference mode.

6.6.2 Continuous presence conferences

When the continuous-presence method of multipoint conferencing is used, each site can see all of the other sites on a continuous basis and switching is not required. This method requires more network resources than with switching. For example, in a four-way conference, each site needs to be able to receive three incoming video signals. Continuous presence may also appear to be more natural to users because the video disruptions caused by switching are avoided.

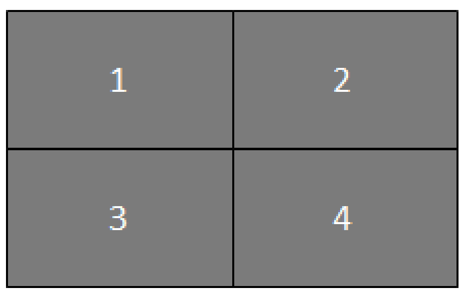

Figure 6.11 Layout 2x2 to display 4 sites continuously

This mode is easily supported by CCNA wherein involved CCEs receive all streams from all users. Streams that flow between CCEs are combined to reduce bandwidth usage at the core. Only one stream is sent from anchor CCE to mobile endpoint.

6.6.3 Lecture mode and round-robin Conferences

This mode uses a layout with a large subpicture showing the lecturer. Video streams of students occupy smaller subpictures. The lecturer subpicture is locked, and the student subpictures operate in continuous presence mode with voice activated priority, so that a student asking a question becomes active in one of the smaller subpictures. There is another mode in which the main image cycles through all the participants over a period of time, it is called round-robin mode

6.7 Mobility Management in CCNA

There are two kinds of mobility one is Inter-CCE mobility which is similar to inter domain mobility that occurs when a mobile endpoint roams across CCEs, and another is Intra-CCE mobility which occurs when a mobile endpoint roams within access networks connected to the same CCE. Below we will explain both of them in detail.

6.7.1 Mobility (Inter-CCE)

While cellular architectures support mobility at core network mainly by using MSC along with information stored in Home Location Registers (HLR), Visitor Location Registers (VLR) and secure information stored in SIM card such as IMSI number, location area identity, secret cryptographic key etc. but in CCNA architecture mobility is supported by using dynamic registration state in network elements setup by ever mobile/roaming end points/devices. Both Intra-CCE and Inter-CCE mobility are simple and straightforward to implement with CCNA. Mobility while roaming under a single CCE (Intra-CCE) is achieved by use of Fixed Mobile techniques/vertical handoffs and switching capabilities of a CCE, we will discuss more about intra-CCE mobility in section 6.7.2. In the case of Inter-CCE when a mobile end-point moves under a new CCE, it sends SIP REGISTER message to the new host CCE which in-turn will establish a new registration state in CCNA, the packets get routed automatically to new location based on new registration state. The new host CCE will communicate with old anchor point CCE and runs load distribution algorithm to determine new target CCE. If target CCE remains the same (old anchor point) then nothing is changed, if the new target CCE is different than the old anchor point then roaming mobile endpoint is notified of new location to send/receive streams. Once target CCE selection and updation occurs the old leg with old host CCE is dropped. The roaming mobile endpoint is in communication with both old host CCE and new host CCE only for the duration of handoff which occurs in a make-before-break fashion. Well-known techniques can also be used to mask mobility for example below the transport layer [58], at the transport layer [56], or at the session layer [57].

6.7.2 Mobility (Intra-CCE) and support for non-SIP mobile endpoints

When a mobile device roams within the access networks connected to a CCE it is called Intra-CCE mobility. The anchor CCE provides mixing and switching capabilities for the mobile endpoints anchored to it, if the mobile-end points are all SIP based, then mobility is provided by technique as specified in N. Banerjee et al [61]. But if a mobile-end point is not SIP Capable or it is on a network which does not provide SIP signaling then we will use techniques of Fixed Mobile Convergence for the anchor CCE to still provide mobility across heterogeneous networks. We will take example of two access networks: WiFi and Cellular (PSTN). Asterisk and uMobility together are chosen to emulate functionalities of CCE because of the existing functionalities such as SIP Gateway, Conference Server, PSTN Gateway, FMC, MCU etc. Whenever a mobile device makes or receives a call, the call is always routed to the CCE. CCE acts as anchor point for both WiFi calls and Cellular Calls, When a user on WiFi places a video call to another device by sending INVITE(P:I) to CCE. CCE handles it by creating a conference session and determines if the called party entry is in its registration table or if it is a customer or forwarded to parent or if it is on WiFi/GSM/Cellular and performs the necessary bridging to add a leg on that particular access network to the conference session previously created. Thereby two users are now connected. For ex. If CCE determines the called party is on GSM then CCE would dial the registered number using one of its ISDN gateway/SIP Gateway. if the caller device moves out of WiFi coverage area, just before WiFi coverage is lost, CCE dials the GSM number of caller to add leg on GSM. Similarly when caller device comes to a WiFi coverage area, WiFi leg is added using SIP connectivity and GSM leg is removed. If the caller and called party are on the same CCE and in Enterprise network, the time taken to create conference session can be saved by directly connecting these intra-enterprise devices and later involving a conference session if user device moves out of WiFi Coverage.

6.7.3 Intra-CCE mobility scenarios

If we assume two mobile endpoints A (Participant 1) and B (Participant 2) under the same CCE, A is a non-SIP phone and can access only a cellular interface, but B is dual-mode phone and capable of accessing both cellular and WiFi.

a) A's cellular interface calls B's cellular interface

In this scenario, the call made by A(Participant 1) reaches CCE through the ISDN gateway, then CCE will create a REGISTER message by generating a temporary mobile-id (security key and id) and establishing a registration state in case if it roams across CCEs. It will check the registration state of B (Participant 2), but since B is on cellular interface CCE will connect user B on cellular interface. Since B is capable of interfacing on WiFi, if B switches to WiFi then CCE will add leg on WiFi interface and only for duration of handoff. Both interfaces of B(cellular, WiFi) and interface A are communicating, but once user B's Wi-Fi leg is added, leg on cellular interface is removed.

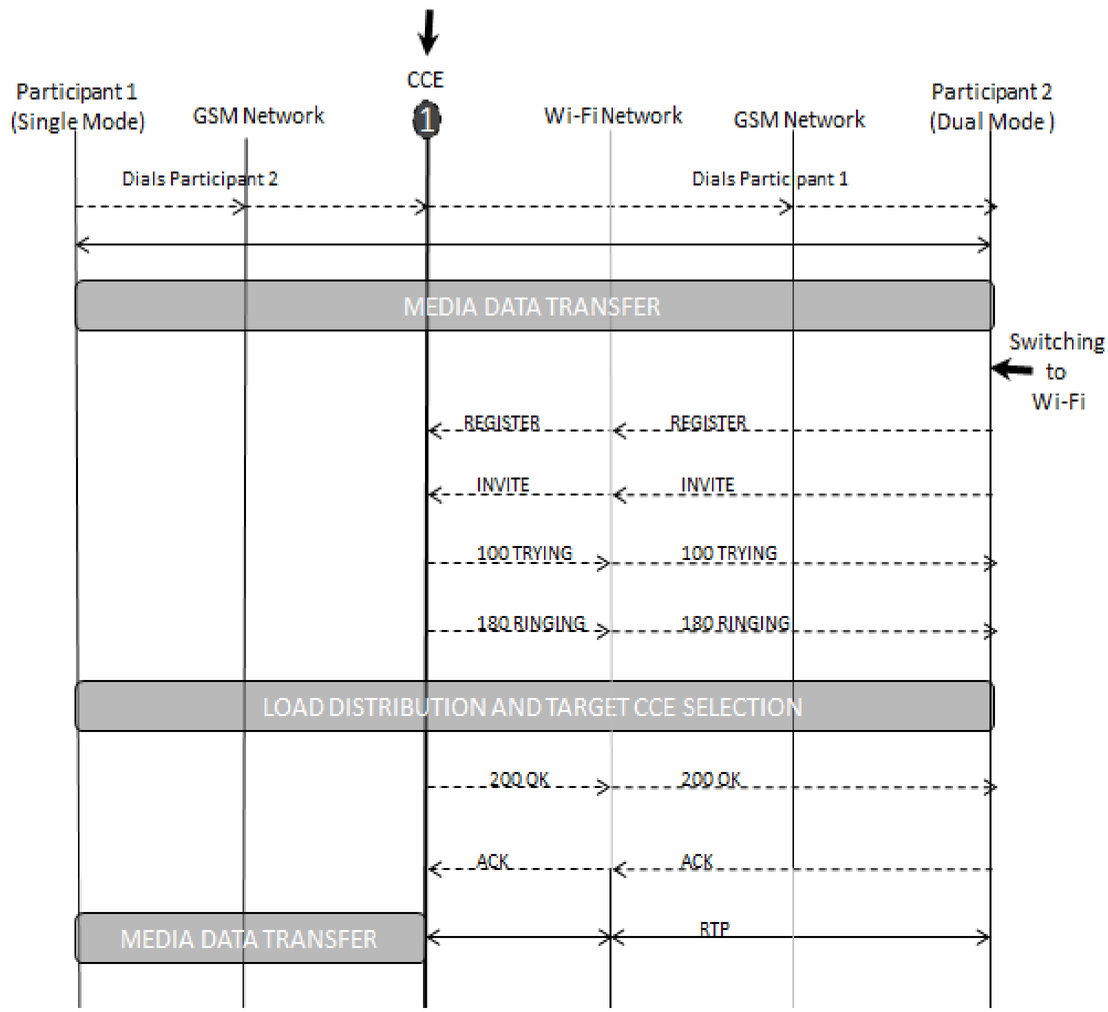

Figure 6.12 Scenario A (Intra-CCE mobility)

Scenario 2: User A's cellular interface calls user B's IP interface

In this scenario, the call made by A (Participant 1) reaches CCE through the ISDN gateway, then CCE will create a REGISTER message by generating a temporary mobile-id (security key and id) and establishing a registration state in case if it roams across CCEs. Then it adds mobile-end point B (Participant 2) which is under the same CCE on Wi-Fi interface. Since B is capable of interfacing on cellular, if B switches to Cellular then CCE will create a temporary mobile-id for it and add leg on cellular interface and only for duration of handoff. Both interfaces of B (cellular, WiFi) and interface A are communicating but once user B's cellular leg is added, leg on Wi-Fi interface is removed.

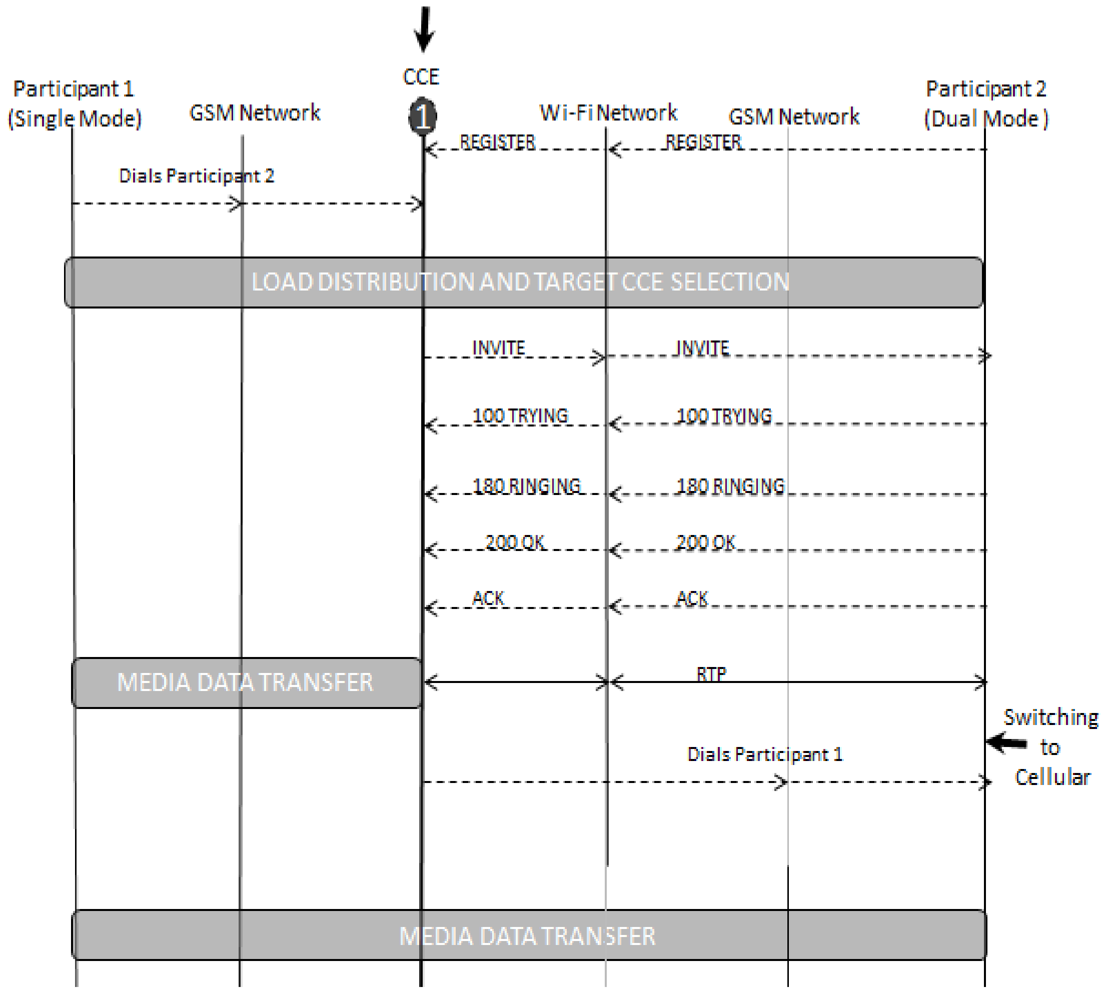

Figure 6.13 Scenario B (Intra-CCE mobility)

Scenario 3: User B's IP interface calls user A's cellular interface

In this scenario, the call made by B reaches CCE, and then CCE will dial A using cellular interface. Since mobile-end point B is capable of interfacing on cellular, if B switches to Cellular then CCE will create a temporary mobile-id for it and add leg on cellular interface and only for duration of handoff. Both interfaces of B(Cellular, WiFi) and interface A are communicating but once user B cellular leg is added, leg on Wi-Fi interface is removed.

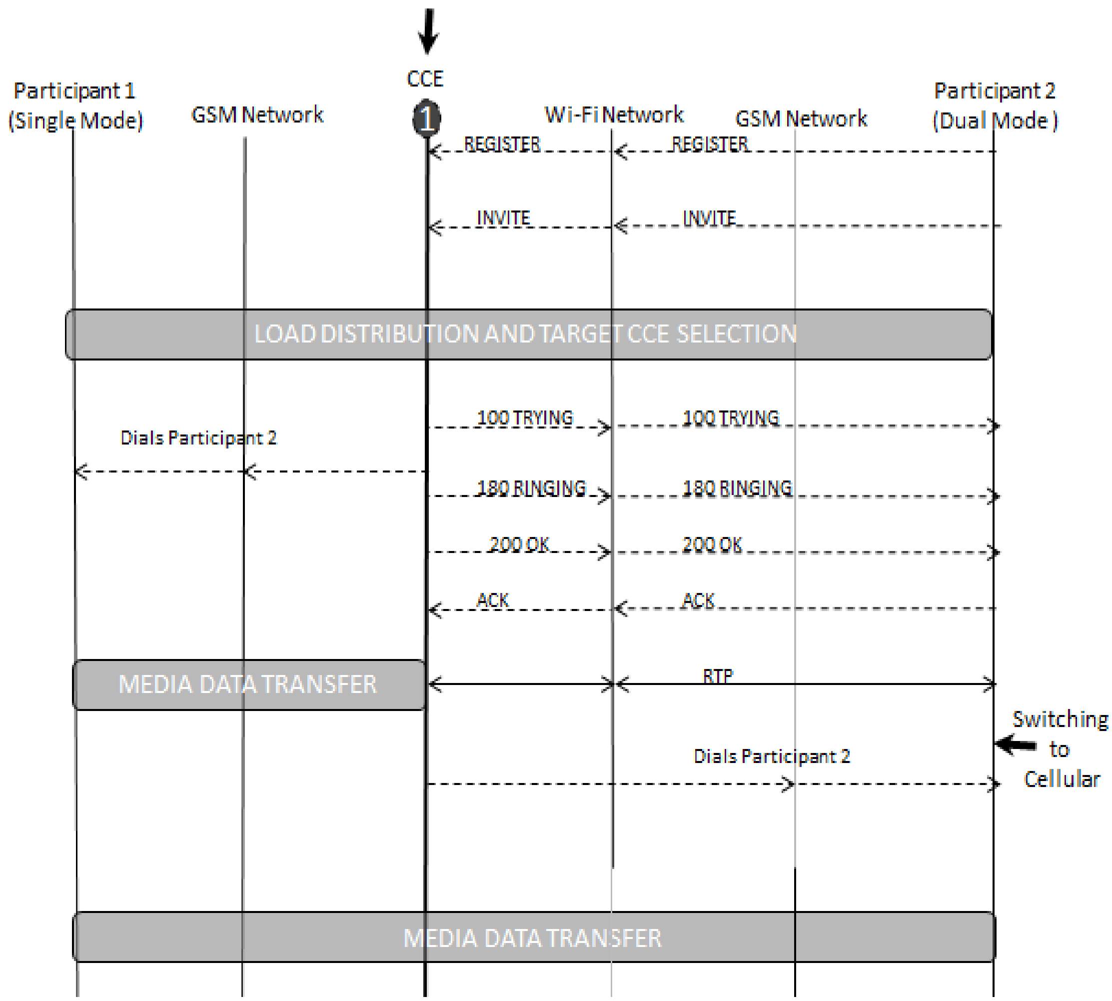

Figure 6.14 Scenario C (Intra-CCE mobility)

Scenario 4: User B's cellular interface calls user A's cellular interface

In this scenario, the call made by A reaches CCE through the ISDN gateway, then CCE will create a REGISTER message by generating a temporary mobile-id (security key and id) and establishing a registration state in case if it roams across CCEs. It will check if user B is registered, but since B is on Cellular Interface, it dials B on cellular interface. Since B is capable of interfacing on WiFi, if B switches to WiFi then CCE will add leg on WiFi interface and only for duration of handoff. Both interfaces of B (Cellular, WiFi) and interface A are communicating but once user B WiFi leg is added, leg on cellular interface is removed.

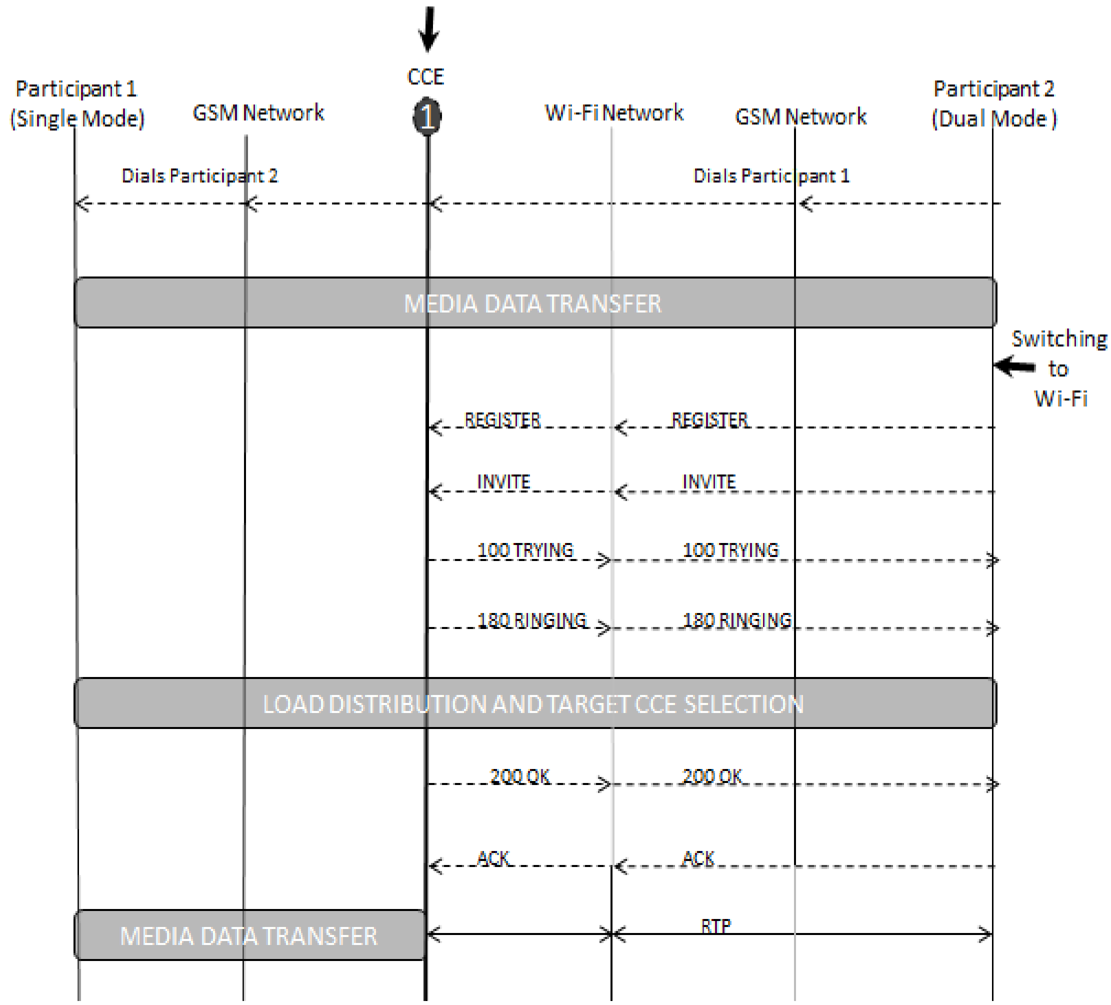

Figure 6.15 Scenario D (Intra-CCE mobility)

6.8 Test Bed and Results

We conducted experiment on voice calls involving dual mode devices by using "uMobility System" provided by Varaha Inc. this shows seamless mobility when mobile roams within access networks connected to a CCE (Intra-CCE mobility).

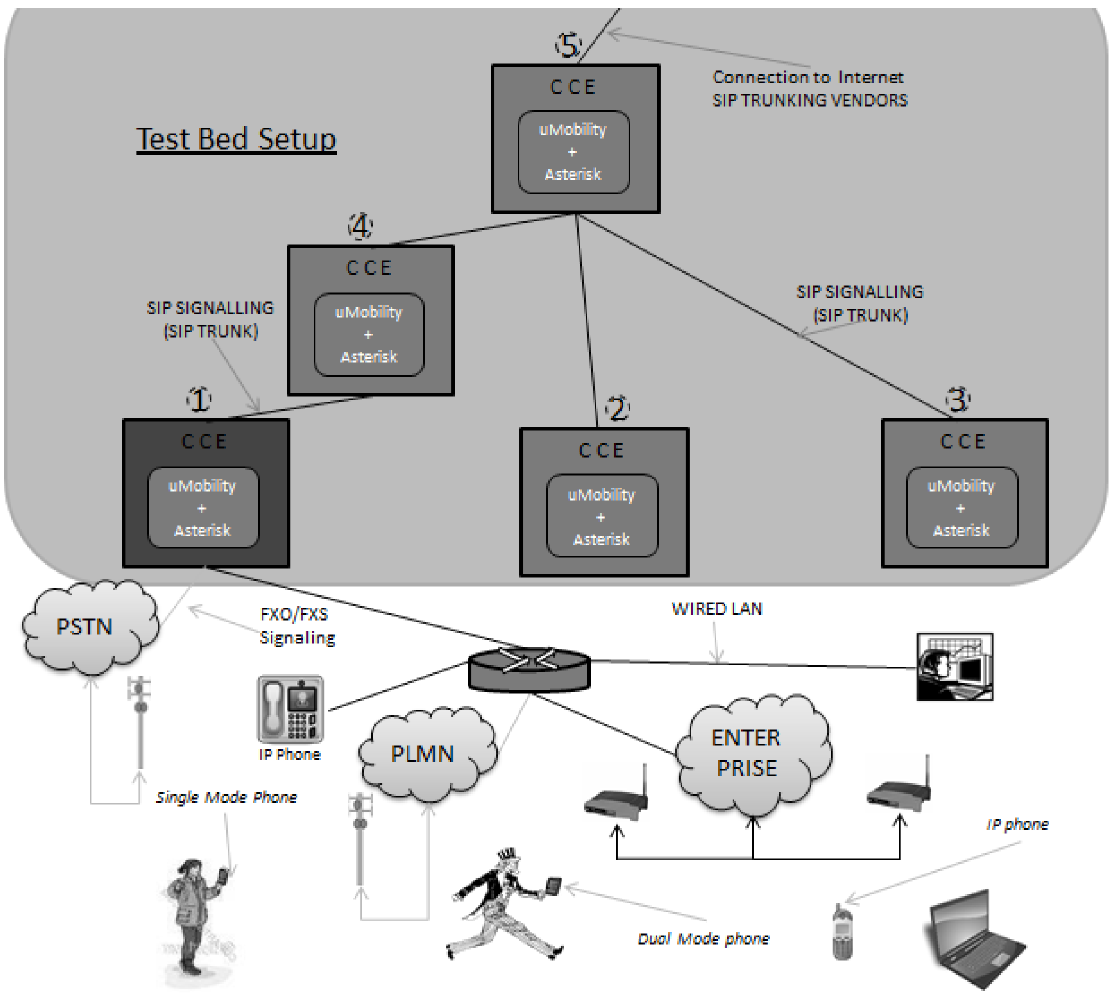

Figure 6.16 Home/Enterprise based FMC

As shown in above figure 6.16, we have set up an home/enterprise based FMC test bed using the uMobility FMC solution by Varaha Systems [21]. The CCE Network element is represented by combined functionalities of uMobility server and Asterisk server. This test bed shows the mobility at edge network where multiple access networks converge. The IP/PBX has been set up using the open source Asterisk PBX software. The uMobility solution comprises of mobility controller and uMobility client. The uMobility controller in conjunction with the asterisk IP/PBX is configured to deliver various FMC services. The uMobility controller based on SIP has the following primary building blocks to achieve FMC.

- Network and Attachment Control Function
- Resource Allocation and Admission Control Function
- Session and Voice Call Continuity Function
- Call Setup, Control and Transfer Function
- Network and User Mobility Management Function
- SIP Proxy, Registrar, Gateway and Redirect server
- Location and Presence Server
- VoIP Gateway

The mobility controller handles voice call continuity as shown in Figure 6.16 The mobility controller acts as a back-2-back user agent (B2BUA) which mediates the call between the call originator (User Agent Server (UAS)) and destination client (User Agent Client (UAC)). It maintains the complete set of calls and connectivity status of UAS/UAC. The IP/PBX is the signaling gateway for the mobility controller. The uMobility client is henceforth referred as UAS or UAC. The IP/PBX is the B2BUA for the mobility controller and does signal conversion for it when initiating, transferring or terminating calls to external IP or wireline network. Registration of UAS/UAC to SIP registrar, i.e. mobility controller, is done using SIP REGISTER method. On successful registration the SIP registrar replies with a 200 OK response message. SIP OPTION method is used by the requesting device to learn the capabilities of the server for call management. SIP SUBSCRIBE method is used by the mobile device and mobility controller to subscribe to event notifications of each other. Event notifications may include feedback on user experience as well as network connectivity to enable mobility controller make intelligent decisions on session and network handover.

The event notifications are sent by SIP NOTIFY method by respective devices. UAS initiates a call using SIP INVITE, which contains the SIP Uniform Resource Identifier (URI) to identify the SIP user agents (UA). For exemplary purpose we assume the IP/PBX uses a SIP trunk to provide connectivity to backhaul network, thus the requirement of the SIP signaling gateway by the trunk provider. Successful call setup is indicated to UAS by “200 OK” response message. After the call setup, the media exchange is done using Real-time Transport Protocol (RTP). Network state can be determined and shared between UAS and mobility controller from feedback through SIP NOTIFY messages as well as periodic SIP REGISTER methods by the UAS to indicate its presence and location in the network. Further the round trip time (RTT) from the two-way handshake of SIP methods can be exploited by the mobility controller to gain a approximation of the delay in the network for session management. As the UAS is moving out of the data network or detects poor quality of experience by the end user it notifies the mobility controller which collaborates with device and network to perform an access network selection.

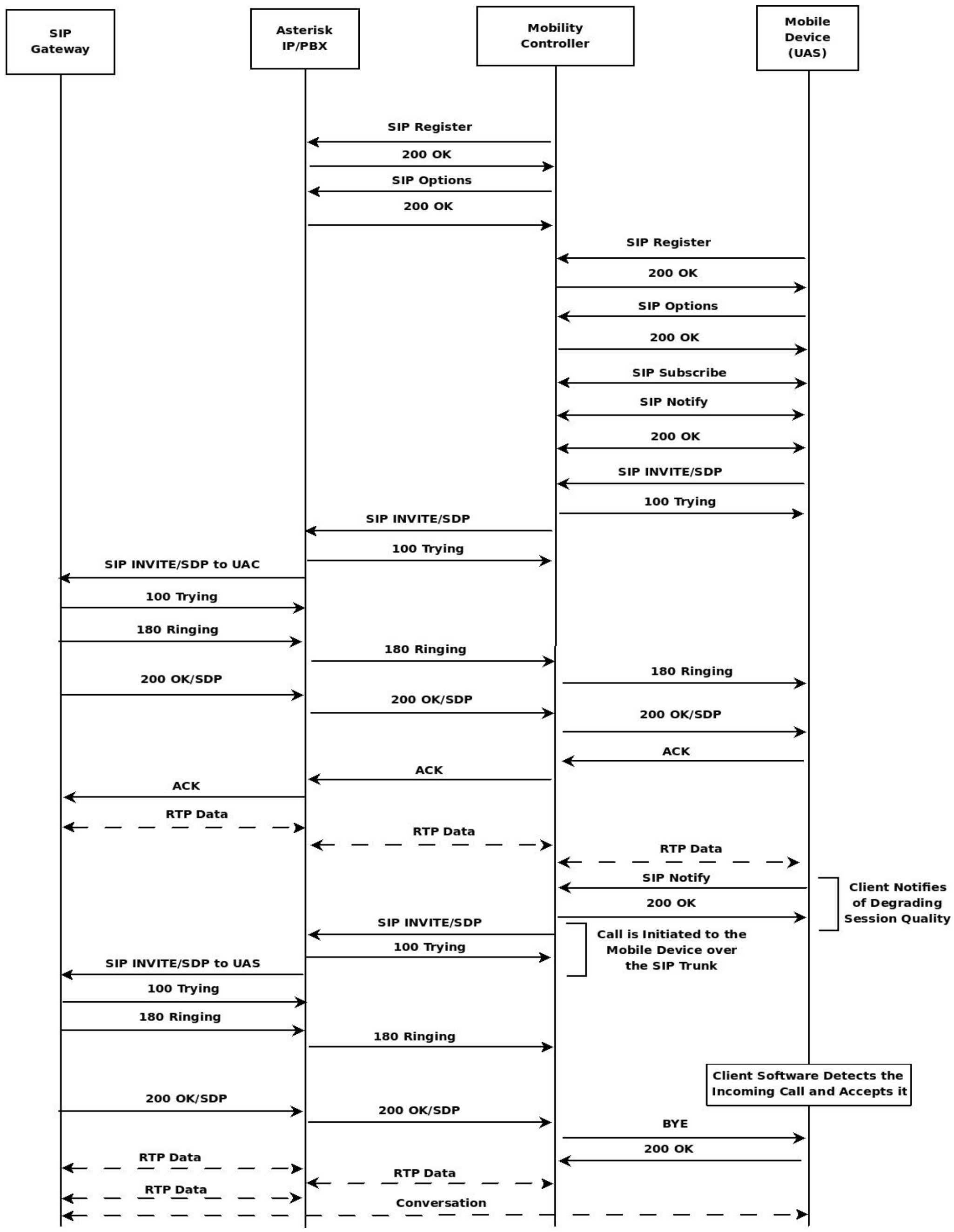

Figure 6.17 Call Flow Graph for Session Transfer

The above figure shows the scenario when the user is moving out of the data network to a GSM network. Before breaking the call, the uMobility controller initiates a call to the GSM number of the UAS through IP/PBX, registered with the mobility controller. The client on UAS detects the incoming call through GSM from mobility controller and accepts it. On successful call setup the uMobility controller redirects the media from over the data network to the GSM through IP/PBX. On redirection and synchronization of new call, the previous call over the wireless data network is terminated with the BYE method. Thus successful call continuity is achieved between wireless data network and GSM with minimal delay. When the UAS/UAE moves back into the wireless data network it registers itself with the mobility controller informing its presence and the call is similarly transferred to it in a “make before break” manner [61]. In the scenario of an incoming call the mobility controller detects through its presence module whether the user is present in the wireless data network and transfers the call to it otherwise redirects the call to its GSM number through IP/PBX. The common Direct Inward Dialing (DID) number used by the mobility controller and IP/PBX from the trunk to identify the UA acts as the single identity of the end user and is used for all practical communication. The mobility client software provides a common UI framework across all mobile devices while assisting the mobility controller in network discovery and selection as well as providing feedback on the quality of experience of the user. Linksys WiFi AP using 802.11b/g protocol deployed as part of the university wireless infrastructure and iPod touch, iPhone and Nokia E61i devices running the uMobility client were used for experiments. Experiments were conducted with voice call session of 330 seconds using a SIP trunk for backhaul connectivity to observe traffic over an all IP network. The uMobility user is a mobile user and is in continuous motion for the duration of call. The G.711 codec was used for delivery of digital voice stream in the IP domain. In the experiments conducted, the traffic was observed at the Asterisk IP/PBX in either direction and average jitter values were observed. The jitter values are averaged over a second, as shown in Figure 6.18.

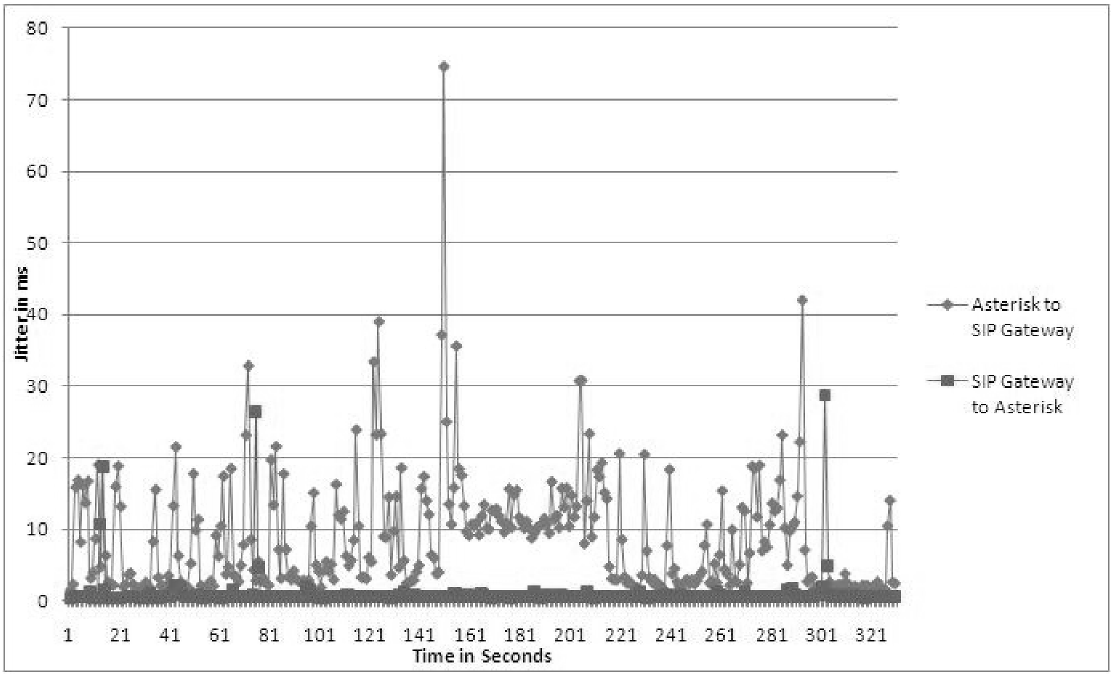

Figure 6.18 Average Jitter of VoIP Traffic in ms

The user moves out of the coverage of the WiFi network at around 150s characterized by the peak value while the session transfers takes place. The session is transferred effectively without any disruption to the GSM network. The constant jitter observed in the interval 150s to 210 s is due to the RTT of RTP data between IP/PBX and uMobility controller carried over LAN. At around 210s the user moves back in the wireless and the call get transferred over the WiFi domain. The variations in jitter observed at other intervals are due to contention with traffic from other sources in the access network as well as movement across access points while that from the signaling gateway through a PSTN network shows fairly uniform characteristics. While this establishes the claim of call continuity it demonstrates the need for a more optimum resource control and provisioning over the access network in the enterprise environment for QoS guarantees in dense traffic scenarios. Thus we demonstrate that seamless mobility can be easily achieved using FMC solutions at edge of internet or anchored CCE where multiple access technologies converge

The below figure 6.19 shows the time taken for callsetup delay, which is the time duration between a user tapping call button on his cellphone and call being received on the other end. We experimented in different scenarios of network load by logging end user client running on iPhone mobile device into various network access points at different locations on the campus and making calls between user on GSM and user on IP. The below call delay was observed with an average delay of 9.8 seconds

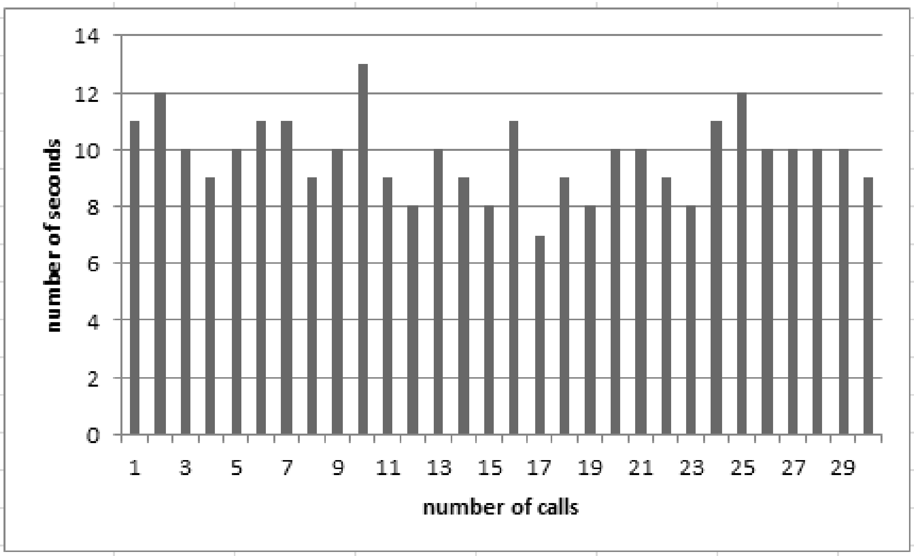

Figure 6.19 Total Call Setup Delay

The below figure 6.20 shows time taken for handoff to occur when mobile endpoint gradually moves to a location where WiFi connection is absent. FMC techniques perform a make-before-break handoff thereby end user is not aware of background transfer of call from IP to GSM. A mobile endpoint triggers handoff with Host CCE based on vertical handoff techniques which includes sensing RF signal strength degradation from anchored access point and Signal to Noise ratio. Handoff delay was calculated based on Asterisk debug logs by measuring time difference between initiation of trigger and receiving ACK from SIP trunk connection to PSTN. Typical handoff delay observed was around .10 seconds during transfer to GSM

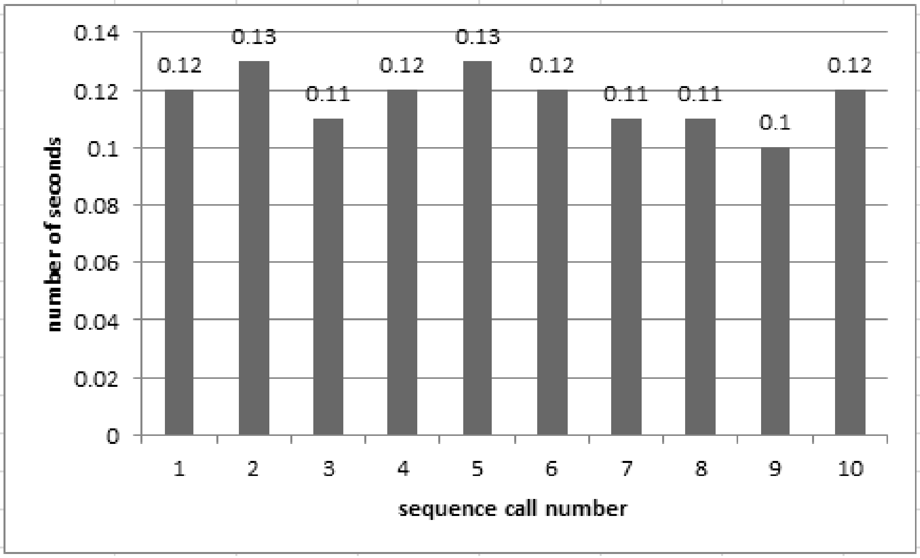

Figure 6.20 Typical Handoff Delay from IP to GSM

The below figure 6.21 shows handoff delay observed during a GSM to IP handoff, which is faster compared with IP to GSM handoff due to higher bandwidth and speed of IP Network. Avergage delay observed was .03 seconds

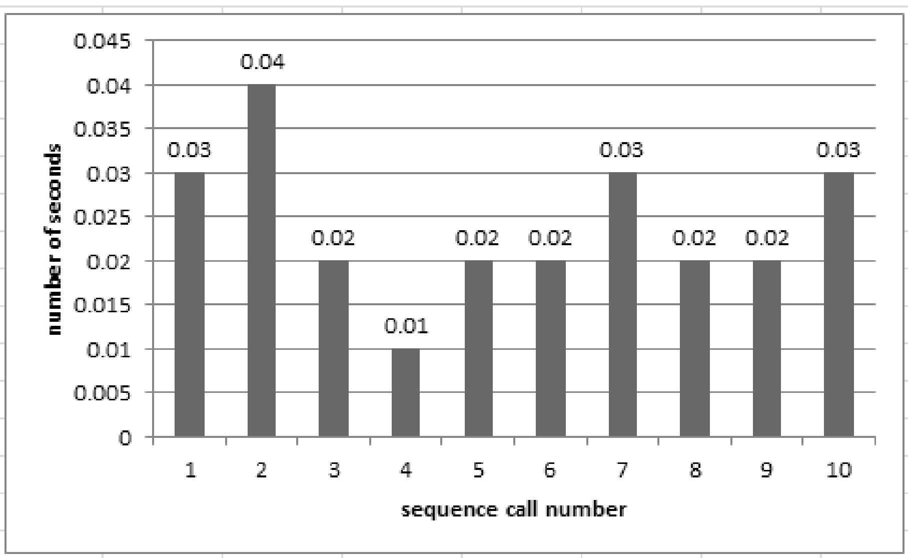

Figure 6.21 Typical Handoff Delay from GSM to IP

CHAPTER 7

CONCLUSION AND FUTURE WORK

In conclusion, we proposed a novel architecture named as Call Control Network Architecture (CCNA) to enable mobility in video conferencing for next generation networks. We introduced a new element named as Call Control Entity (CCE) whose functionalities are derived from popular future telephony tools such as SIP proxy, SIP registrar, Conference server, ISDN gateway and SIP Gateway, available in open source software tools such as Asterisk, OpenSIPS and others, along with newly added functionalities of mobility management, load distribution. We also introduced a new way of mobile identification and security. We described how basic conferencing operations and different conferencing modes supported. We showed how mobility is handled at the core (Inter-CCE) and edge network (Intra-CCE) elements of this architecture. We also setup a test bed based to demonstrate seamless mobility in Intra-CCE scenarios and conducted experiments. We measured performance statistics (such as call setup time, handoff time, jitter) to conclude that these are acceptable based on requirements for VoIP. In our future work we will setup a test bed to emulate the core network involving multiple CCE network elements and conduct experiments

REFERENCES

[1] H.323 Forum - http://www.h323forum.org/

[2] RFC: Session Initiation Protocol - http://www.ietf.org/rfc/rfc3261.txt

[3] RTP: A Transport Protocol for Real-Time Applications - http://tools.ietf.org/html/rfc3550

[4] O'Hara, K., Black, A., and Lipson, M. Media Spaces and Mobile Video Telephony, Media Space: 20+ Years of Mediated Life, Springer (2009), 303-323.

[5] Yi Cui, Klara Nahrsetedt, Dongyan Xu, "Seamless Userlevel Handoff in Ubiquitous Multimedia Service Delivery", Multimedia Tools and Applications Journal (ACM/Kluwer), Special Issue on Mobile Multimedia and Communications and m-Commerce, February 2004.

[6] Henning Schulzrinne and Elin Wedlund, "Application-Layer Mobility Using SIP", Mobile Computing and Communications Review, Volume 4, Number 3, July 2000, pp.47-57.

[7] Henry Song, Hao-hua Chu, Shoji Kurakake, "Browser Session Preservation and Migration", Poster Session of WWW Conference, Hawaii, U.S.A., 7-11 May 2002.

[8] R. Shacham et al, "Session Initiation Protocol (SIP) Session Mobility ", IETF draft-shacham-sipping-session-mobility-02, February 2006.

[9] Rosenberg, J., Peterson, J., Schulzrinne, H., and G. Camarillo, "Best Current Practices for Third Party Call Control (3pcc) in the Session Initiation Protocol (SIP)", RFC 3725, April 2004.

[10] OpenWengo Freedom to Call, Freedom to Code - http://www.openwengo.org/

[11] Columbia SIP user agent - http://www1.cs.columbia.edu/ xiaotaow/sipc/

[12] Windows mobile - http://www.microsoft.com/windowsmobile

[13] Symbian OS - http://www.symbian.com/

[14] Free SIP/VOIP client for Android - http://sipdroid.org/

[15] Apple iPhone FaceTime Calling - http://www.apple.com/iphone/features/facetime.html

[16] End-to-end video quality analysis and modeling for video streaming over IP network Zhihai He; Chang Wen Chen; Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference

[17] Progressive video transmission over a wired-to-wireless network Anand, R.; Podilchuk, C.; Lou, H.; Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE

[18] A Framework for Conferencing with the Session Initiation Protocol (SIP) (http://www.rfc-archive.org/getrfc.php?rfc=4353)

[19] M. Gritter and D. R. Cheriton. TRIAD: A New Next-Generation Internet Architecture. http://www-dsg.stanford.edu/triad/, July 2000.

[20] IEEE 802.21 - www.ieee802.org/21/

[21] Varaha Systems - http://www.varaha.com/

[22] Lo, N. 2006. Corporate communication goes wireless, (Nov. 2006). http://www.enterpriseinnovation.net .

[23] Technical paper, 2005. Better Enterprise Applications on Any Network: A New Era of Intelligent Communications, Avaya, EF-MIS2729.

[24] Schwartz, D. 2005. Why wireline carriers will be the early adopters of fixed mobile convergence, LongBoard, TelephonyOnline, (Feb. 18, 2005).

[25] Brannstrom, R., Kodikara, E.R., Ahlund, C., and Zaslavsky, 2006. A. Mobility management for multiple diverse applications in heterogeneous wireless networks, in Proc. 3rd IEEE Consumer Communications and Networking Conference - CCNC'06.

[26] White paper, 2006. Managing the Mobile Enterprise – A Guide for Creating Effective Enterprise Wireless Strategies, Good Technologies.

[27] Technical Report, 2006. Mobile Enterprise Network Services", an IDC Continuous Intelligence Service, WT_IDC_P4426_0706.

[28] Xiao, Y., et al., 2005. Architecture, mobility management, and quality of service for integrated 3G and WLAN networks, Wiley Wireless Communications

[29] Technical paper, 2006. Extending the Enterprise Network Through Mobility, ProCurve Networking by HP, Hewlett- Packard Development Company, L.P. 4AA0-0542ENW, (4/2006).

[30] Fixed Mobile Convergence (Cisco), http://www.cisco.com/en/US/netsol/n s519/networking solutions solution category.html

[31] A Sheneyderman and A. Casati, "Fixed Mobile Convergence: Voice over WiFi,IMS, UMA/GAN, Femtocells and other Enablers," McGraw-Hill Communications, 2008.

[32] Report on Fixed-Mobile Convergence: Implications on Competition and Regulatory Aspect, European Regulatory Group, March 2009.

[33] T. Pardijs, "Fixed Mobile Convergence: Regulatory Implications of Convergence of Fixed and Mobile Telephony," Master Thesis, OPTA, TU Delft, November 2005.

[34] M. Vrdoljak, S. I. Vrdoljak and G. Skugor, "Fixed-Mobile Convergence Strategy: Technologies and Market Opportunity," IEEE Communications Magazine, vol. 38, Issue 2, pp. 116-121, February 2000.

[35] D. Kataria and F. Logothetis, "Fixed Mobile Convergence: Network Architecture, Services, Terminals and Traffic Management," Proc. of PIMRC, vol. 4, pp. 2289-2300, September 2005.

[36] V. Chandrasekhar, J. G. Andrews and A. Gatherer, "Femtocell Networks: A Survey," IEEE Communications Magazine, vol. 46, no. 9, pp. 59-67, September 2008.

[37] W. Simpson, "Video over IP, A Practical Guide to Technology and Applications," Focal Press Media Technology Professional, 2006.

[38] Fixed-Mobile Convergence: An Integrated Operator Case Study Renjish Kumar, K.R.; Rokkas, T.; Varoutas, D.; Kind, M.; Von Hugo, D.; Harno, J.; Smura, T.; Heikkinen, M.; Telecommunication Techno-Economics, 2007. CTTE 2007. 6th Conference

[39] 3GPP Long Term Evolution - http://www.3gpp.org/LTE

[40] IEEE 802.20 - http://www.ieee802.org/20/

[41] IEEE 802.16 - http://www.ieee802.org/16/

[42] Video support in Asterisk. Asterisk.org. http://www.asterisk.org/doxygen/trunk/AstVideo.html

[43] Md. Zaidul Alam, Saugata Bose, Md. Mhafuzur Rahman, Mohammad Abdullah Al-Mumin, “Small office PBX using Voice over IP” in ICACT I2-14 FEB,2007

[44] Ryosuke Yamamoto,,Fumikazu Iseki,Moo Wan Kim, “Validation of Voip system for University Network” in ICACT,feb 2008.

[45] Asterisk.org, "Features and Architecture of Asterisk PBX", http://www.asterisk.org/features, accessed in March, 2006.

[46] McNair, J., "Handoff Techniques for Next Generation Wireless Multimedia Systems," Georgia Institute of Technology, Atlanta, Georgia, USA, December 2000.

[47] E. S. El-Alfy, Y. D.Yao, H. Heffes, “Autonomous call admission control with prioritized handoff in cellular networks," IEEE Int. Conf. Comm., vol.5, p.1386-1390, June 2001.

[48] A Vertical Handoff Decision Algorithm for Next Generation Wireless Networks Nkansah-Gyekye, Y.; Agbinya, J.I.; Broadband Communications, Information Technology & Biomedical Applications, 2008 Third International Conference

[49] "Fixed Mobile Convergence and Beyond", Richard Watson, 2009

[50] Video Over IP Networks: Subjective Assessment of Packet Loss Kukhmay, Y.; Glasman, K.; Peregudov, A.; Logunov, A.; Consumer Electronics, 2006. ISCE '06. 2006 IEEE Tenth International Symposium

[51] Multipoint conferencing for mobile communications network Li, Y.; Andresen, S.; Universal Personal Communications, 1993. Personal Communications: Gateway to the 21st Century. Conference Record., 2nd International Conference

[52] The IP multimedia subsystem with an e-learning management system and integrated video conferencing Michael Massoth, Roman Korn November 2008 iiWAS '08: Proceedings of the 10th International Conference on Information Integration and Web-based Apps & Services

[53] C. Ziegler and G. Weiss, “Multimedia Conferencing on Local Area Networks”, IEEE Computer, vol. 23, no. 9, pp. 52- 61, September 1990.

[54] Keller and H. Reiser, “Dynamic Management of Internet Telephony Service: A case Study based on JavaBeans and JDMK”, Proceeding of Enterprise Distribute Object Computing Conference, 1999.

[55] K. Singh, G. Nair and H. Schulzrinne, "Centralized Conferencing using SIP", Proceedings of the 2nd IP-Telephony Workshop, April 2001

[56] Distributed management architecture for multimedia conferencing using SIP, Yeong-Hun Cho; Moon-Sang Jeong; Jong-Tae Park; Wee-Hyuk Lee; Distributed Frameworks for Multimedia Applications, 2005. DFMA '05. First International Conference

[57] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, and Ion Stoica. A data-oriented (and beyond) network architecture. In SIGCOMM ’07: Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications, pages 181–192, New York, NY, USA, 2007. ACM.

[58] R. Moskowitz and P. Nikander. Host Identity Protocol Architecture. RFC 4423, IETF, May 2006.

[59] D. Funato, K. Yasuda, and H. Tokuda. TCP-R: TCP Mobility Support for Continuous Operation. In Proc. of ICNP ’97, pages 229–236, Atlanta, GA, USA, 1997.

[60] M. A. C. Snoeren. A Session-based Architecture for Internet Mobility. PhD thesis, MIT, Department of Electrical Engineering and Computer Science, Feb. 2003.

[61] N. Banerjee, A, Acharya, S. K. Das, “Seamless SIP-based Mobility for Multimedia Application,” IEEE Network Magazine, vol. 20, Issue 2, pp. 6-13, April 2006.

[62] K. Singh, G. Nair, and H. Schulzrinne, “Centralized conferencing using SIP,” in Proc. 2nd IP-Telephony Workshop, Apr. 2001, pp. 31–37.

[63] R. V. Prasad, R. Hurni, and H. S. Jamadagni, “A scalable distributed VoIP conferencing using SIP,” in Proc. 8th IEEE Int. Symp. Comput. Commun., Jun./Jul. 2003, pp. 608–613.

[64] N. C. Damianou, “A policy framework for management of distributed systems,” Ph.D. dissertation, Imperial College of Sci., Technol. Med., Univ. London, London, U.K., 2002

[65] An Architectural Framework for Media Server Control - http://www.rfc-archive.org/getrfc.ph p?rfc=5567

BIOGRAPHICAL INFORMATION

Avinash Narayan received his Bachelor of Engineering Degree in Computer Science and Engineering from Rashtreeya Vidyalaya College of Engineering, Bangalore, India, in 2003. He Worked for Motorola, India for 4.5 years in Call Processing Development Team of “Mobility Manager” Product for 3G CDMA 1X wireless infrastructure. He also worked as a Research Assistant in Indian Institute of Science, Bangalore for 1 year. He began his graduate studies in the Department of Computer Science and Engineering at The University of Texas at Arlington in Fall 2008. He received his Master of Science degree in Computer Science and Engineering from The University of Texas at Arlington in August 2010. His current research interests are Fixed Mobile Convergence, Next Generation Network Architecture, Mobility Management, Multiagent Systems, Algorithm design, Nash Equilibrium and Game theory.