

國立交通大學

資訊工程學系

博士論文

網路多媒體服務之通話控制

Call Control for IP Multimedia Service



研究生：許孟達

指導教授：林一平 博士

張明峰 博士

中華民國九十六年一月

網路多媒體服務之通話控制

Call Control for IP Multimedia Service

研究生：許孟達

Student : Meng-Ta Hsu

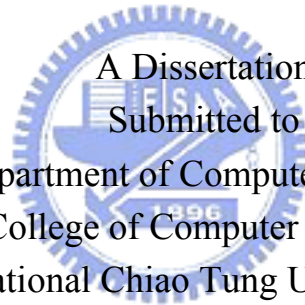
指導教授：林一平 博士

Advisor : Dr. Yi-Bing Lin

張明峰 博士

Dr. Ming-Feng Chang

國立交通大學
資訊工程學系
博士論文



A Dissertation
Submitted to
Department of Computer Science
College of Computer Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Computer Science

January 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年一月

國立交通大學

博碩士論文全文電子檔著作權授權書

(提供授權人裝訂於紙本論文書名頁之次頁用)

本授權書所授權之學位論文，為本人於國立交通大學 資訊工程 系所 網路工程 組，95 學年度第 1 學期取得博士學位之論文。

論文題目：網路多媒體服務之通話控制

指導教授：林一平博士，張明峰博士

同意 不同意

本人茲將本著作，以非專屬、無償授權國立交通大學與台灣聯合大學系統圖書館：基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學及台灣聯合大學系統圖書館得不限地域、時間與次數，以紙本、光碟或數位化等各種方法收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行線上檢索、閱覽、下載或列印。

論文全文上載網路公開之範圍及時間：

本校及台灣聯合大學系統區域網路	<input checked="" type="checkbox"/> 中華民國 96 年 1 月 15 日公開
校外網際網路	<input checked="" type="checkbox"/> 中華民國 96 年 1 月 15 日公開

授權人：許孟達

親筆簽名：許孟達

中華民國 96 年 1 月 15 日

國立交通大學

博碩士紙本論文著作權授權書

(提供授權人裝訂於全文電子檔授權書之次頁用)

本授權書所授權之學位論文，為本人於國立交通大學 資訊工程 系所 網路工程 組，95 學年度第 1 學期取得碩士學位之論文。

論文題目：網路多媒體服務之通話控制

指導教授：林一平博士，張明峰博士

■ 同意

本人茲將本著作，以非專屬、無償授權國立交通大學，基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學圖書館得以紙本收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行閱覽或列印。

本論文為本人向經濟部智慧局申請專利(未申請者本條款請不予理會)的附件之一，申請文號為：_____，請將論文延至____年____月____日再公開。

授權人：許孟達

親筆簽名： 許孟達

中華民國 96 年 1 月 15 日

國家圖書館博碩士論文電子檔案上網授權書

ID:GT008917587

本授權書所授權之論文為授權人在國立交通大學 資訊學院 資訊工程 系所 網路工程 組 95 學年度第 1 學期取得碩士學位之論文。

論文題目：網路多媒體服務之通話控制

指導教授：林一平博士，張明峰博士

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

※ 讀者基於非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：許孟達

親筆簽名：許孟達

民國 96 年 1 月 15 日

1. 本授權書請以黑筆撰寫，並列印二份，其中一份影印裝訂於附錄三之二(博碩士紙本論文著作權授權書)之次頁；另一份於辦理離校時繳交給系所助理，由圖書館彙總寄交國家圖書館。

國立交通大學
資訊工程系博士班

論文口試委員會審定書

本校 資 訊 工 程 系 許孟達 君

所提論文 網路多媒體服務之通話控制

合於博士資格水準、業經本委員會評審認可。

口試委員： _____

指導教授： _____ _____

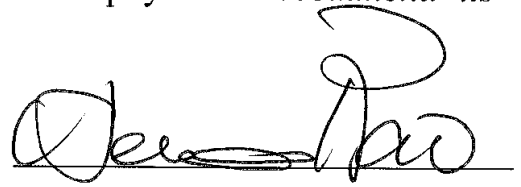
系主任： _____

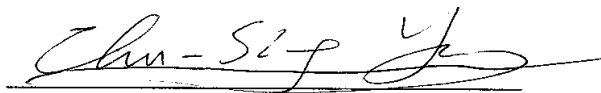
中華民國九十六年一月九日

Department of Computer Science
College of Computer Science
National Chiao Tung University
Hsinchu, Taiwan, R.O.C.

Date: January 9, 2007

We have carefully read the dissertation entitled Call Control for IP Multimedia Service submitted by Meng-Ta Hsu in partial fulfillment of the requirements of the degree of Doctor of Philosophy and recommend its acceptance.



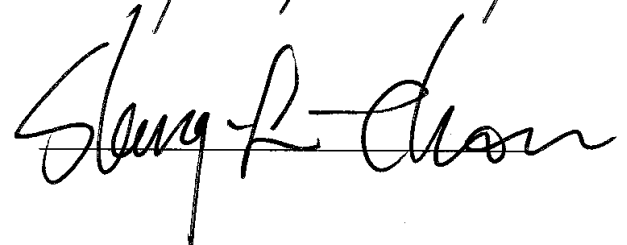




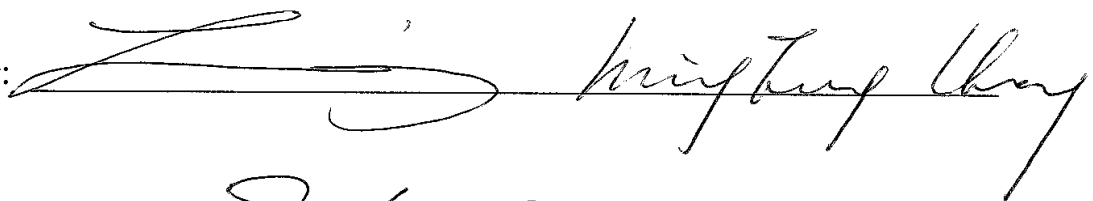




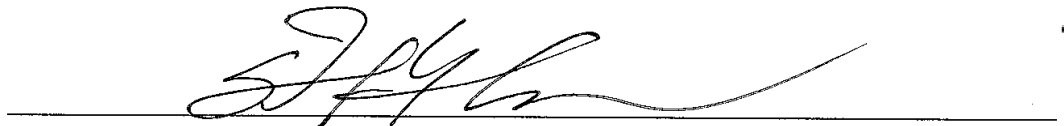




Thesis Advisor:



Chairman:



網路多媒體服務之通話控制

學生：許孟達

指導教授：林一平 博士
張明峰 博士

國立交通大學資訊工程學系

摘 要

這幾年來，有很多的力量正驅使著系統供應商提供整合的服務。第一，消費者希望在各式各樣的機器上面可以享受到統一的服務。第二，消費者希望可以不管使用的是哪一種接取服務，不管是 WiFi，3G，Ethernet，WiMax，ADSL 或 CABLE，都能享受到同樣的服務內容。第三，有線系統業者的利潤正在降低，因此他們必須開始在無線的領域尋找機會，而無線系統業者的成長速度減慢，也使他們開始探索如何擴大市場。最後，無線與有線業者之間的競爭將會驅使網路服務邁向整合，而網際網路上的某些語音服務的成功案例，如 Yahoo! Messenger 和 Skype 等，則使得網際網路應用在網路服務邁向整合的過程中佔了舉足輕重的角色。

由於使用者對行動通訊需求越來越高，他們希望在任何地方都能無線的連接 Internet。3GPP 所定義的 IP Multimedia Subsystem (IMS) 因此而變得更加重要。本論文中，我們分別就網路多媒體服務的各層面討論設計的議題。在網路電話的整合網路中，我們提出了一個整合的架構稱為 Integrated Call Agent (iCA)。iCA 整合了網路通訊中的 Media Gateway Control Protocol (MGCP)、Session Initiation Protocol (SIP) 與 H.323 三個主要的協定，讓使用者可以在整合的服務中沒有障礙的互通。我們參考 Intelligent Network (IN) 並提出了很彈性的架構，讓將來整合進更多協定成為可能並且所花的心力降到最低。

針對使用者的身分認證問題，3GPP 制定了 two-pass authentication 程序，分別為 General Packet Radio Service (GPRS) 網路以及 IMS 網路認證使用者。我們發現在 two-pass authentication 中有許多重覆的步驟，因此在論文中我們

提出了 one-pass authentication 的方法，以 GPRS 網路執行相同的認證程序，但 IMS 網路以簡化的方法在使用者註冊過程完成認證。我們證明 one-pass authentication 可以正確的認證使用者，同時省下 50%的訊息交換。

在網路多媒體服務中的服務平台中，我們探討了 SIP 基礎的網路電話客服中心的設計與實作，並且提出了等待時間的演算法。我們設計了兩個評量預測準確度的方法並且發展了一個模擬模型來測量演算法的準確度。我們發現所提出的演算法可以有效的控制使用者等待超時的機率。

以上的研究成果提供讀者在研究網路多媒體服務的通話控制以及身分認證的議題上，可供參考之基礎。

關鍵字: 第三代行動通訊、Universal Mobile Telecommunications System (UMTS)、多媒體子系統、身分認證、Session Initiation Protocol (SIP)、通話控制、Media Gateway Control Protocol (MGCP)、等待時間預測。



Call Control for IP Multimedia Service

Student: Meng-Ta Hsu

Advisors: Dr. Yi-Bing Lin

Dr. Ming-Feng Chang

Department of Computer Science
National Chiao Tung University

ABSTRACT

In recent years, there are many forces urging operators to provide the converged services. First, customers hope to enjoy unified services on different type of devices. Second, users want to have uniform services and contents regardless of the kind of underlying access network being used, such as WiFi, 3G, ethernet, WiMax, ADSL, or cable. Third, fixed network providers' profit is reducing, so they must begin to look for the opportunities in wireless domain. And wireless operators' speed of growth slows down, makes them begin to explore how to expand their markets. Finally, the competition between the fixed and wireless carriers will drive the network service toward convergence, and some successful cases of Internet telecommunications services such as Yahoo! Messenger and Skype, etc., make Internet Protocol (IP) applications take very important role during the process toward convergence.

As consumers become increasingly mobile, they will demand wireless Internet access from everywhere. In keeping with these requirements of end users, IP Multimedia Subsystem (IMS) standards based on the 3GPP UMTS become more and more important. In the IMS system, we propose an integrated call agent of the converged VoIP network. We present a simple, flexible framework for the interworking functions of VoIP protocols such as Media Gateway Control Protocol (MGCP), Session Initiation Protocol (SIP) and H.323 base on Intelligent Network (IN).

In UMTS two-pass authentication, many steps in the General Packet Radio Service (GPRS) authentication and IMS authentication are duplicated. Therefore, we propose an one-pass authentication procedure, in which only the GPRS authentication procedure is performed. In the IMS network, the authentication is implicitly executed in the IMS registration. We formally prove that the IMS user is correctly authenticated, and the one-pass authentication saves up to 50% of the IMS registration/authentication traffic.

In the service platform of IP multimedia services, we describe the design and implementation of a SIP-based VoIP call center with waiting time prediction. The SIP-based plug-in modular call center architecture and detailed message flows are elaborated. We propose two output measures and develop a discrete event simulation model to investigate the performance of the waiting time prediction algorithm for the call center.

These research results presented in this dissertation can be viewed as a useful foundation for further study in call control for IP multimedia services and authentication.

Key Words: Third Generation (3G), Universal Mobile Telecommunications System (UMTS), IP Multimedia Subsystem (IMS), authentication, Session Initiation Protocol (SIP), call control, Media Gateway Control Protocol (MGCP), waiting time prediction, Call Session Control Function (CSCF).

Acknowledgement

I would like to express my sincere thanks to my advisors, Prof. Yi-Bing Lin and Prof. Ming-Feng Chang. Without their supervision and perspicacious advice, I can not complete this dissertation. Special thanks to my committee members, Dr. Sheng-Lin Chou, Prof. Han-Chieh Chao, Dr. Herman Chung-Hwa Rao, Prof. Chu-Sing Yang, Prof. Wen-Nung Tsai, and Prof. Hsi-Lu Chao for their valuable comments. Thanks also to the colleagues in Laboratory 117 and 118.

I also express my appreciation to all the faculty, staff and colleagues in the Department of Computer Science and Information Engineering. In particular, I would like to thank Prof. Phone Lin, Prof. Ai-Chun Pang, Dr. Yuan-Kai Chen, Prof. Wei-Zu Yang, Prof. Shun-Ren Yang, Prof. Pei-Chun Lee for their friendship and support in various ways.

Finally, I am grateful to my family, my dear father, mother, sister, Joan Hsu, my little baby, and friends for their encouragement and support during these years.

Contents

Abstract in Chinese	i
Abstract in English	iii
Acknowledgement	v
Contents	vi
List of Tables	ix
List of Figures	x
Abbreviation	xii
1 Introduction	1
1.1 IP Multimedia Subsystem	2
1.2 An Integrated Call Agent of the Converged VoIP Network	7
1.3 One-Pass GPRS and IMS Authentication Procedure for UMTS	8
1.4 A SIP-based Call Center with Waiting Time Prediction	10
2 An Integrated Call Agent of the Converged VoIP Network	12
2.1 Introduction	13
2.1.1 VoIP protocols	13
2.1.2 Interoperation	16
2.2 IN Basic Call State Model	17
2.3 System Architecture	19



2.3.1 Mapping of VoIP protocol messages to the BCSM messages	22
2.3.2 H.323 slow-start	25
2.4 System Implementation and Result	27
2.5 Conclusions	29
3 One-Pass GPRS and IMS Authentication Procedure for UMTS	31
3.1 Introduction	31
3.2 3GPP Two-Pass Authentication	35
3.3 One-Pass Authentication Procedure	41
3.4 Correctness of the One-Pass Procedure	46
3.5 Summary	50
4 A SIP-based Call Center with Waiting Time Prediction	51
4.1 Introduction	51
4.2 Automatic Call distributor Module for SER	53
4.3 Call Center Message Flows	56
4.4 The Waiting Time Prediction Algorithm	63
4.4.1 Enhanced Whitt's Algorithm	64
4.4.2 Performance Evaluation	66
4.5 CallBack Mechanism	69
4.5.1 The PTN CallBack Mechanism	71
4.5.2 Discussions	72
4.6 Conclusions	73
5 Conclusions and Future Work	74
5.1 Summary	74
5.2 Future Works	75

Reference 77

Curriculum Vita 82

Publication List 83



List of Tables

3.1 Identical Steps in GPRS and IMS Authentications	39
3.2 Comparing the One-Pass and the Two-Pass Authentication Procedures in IMS Registration	43



List of Figures

1.1 UMTS All-IP Network Architecture	4
2.1 MGCP architecture	16
2.2 IN architecture	18
2.3 Simplified IN BCSMs	19
2.4 Components developed for a general VoIP gateway	20
2.5 A SIP/H.323 gateway	20
2.6 A converged VoIP network using gateways	21
2.7 A converged VoIP network managed by integrated call agents	22
2.8 An example of H.323 and MGCP interworking using 2 ICAs	22
2.9 Mapping VoIP messages to BCSM messages	24
2.10 BCSMs for the H.323 slow-start	26
2.11 Call flow of H.323 (slow-start) and SIP interworking	27
2.12 Components used in our platform	29
2.13 Call establishment delays	29
3.1 UMTS architecture for packet switched service domain	33
3.2 Message flow for 3GPP GPRS authentication	35
3.3 Message flow for 3GPP IMS authentication	37
3.4 Illegal IMS registration	40
3.5 IMS registration (one-pass authentication)	42
3.6 Improvement of the One-Pass Procedure over the Two-Pass Procedure	46

4.1 VoIP-based call center	52
4.2 The proposed call center architecture	54
4.3 The FSM State Transition Diagram of ACD	56
4.4 The select_agent() function	58
4.5 The set_agent_state() function	59
4.6 Normal call setup message flow	60
4.7 The flow diagram for the agent dispatcher	61
4.8 Message flow when all agents are busy	62
4.9 The δ and θ performances ($n=80, \mu=0.05/\text{sec}, \gamma=0.01$)	67
4.10 The δ and θ performances when customers may be impatient	69
4.11 Private Telecommunications Network Architecture	70



Abbreviation

The abbreviations used in this dissertation are listed below.

3GPP: 3rd Generation Partnership Project

ACD: Automatic Call Distributor

AKA: Authentication and Key Agreement

ALG: Application Level Gateway

AuC: Authentication Center

BCSM: Basic Call State Model

CA: Call Agent

CS: Circuit-Switched

CSCF: Call Session Control Function

DTMF: Dual Tone Multi-Frequency

FSM: Finite State Machine

GGSN: Gateway GPRS Support Node

GMM: GPRS Mobility Management

GPRS: General Packet Radio Service

GSM: Global System for Mobile Communications

HLR: Home Location Register

HSS: Home Subscriber Server

ICA: Integrated Call Agent

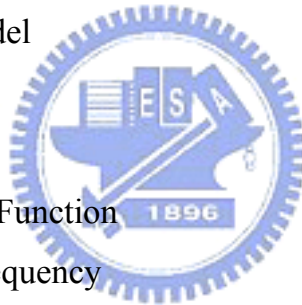
IMPI: IP Multimedia Private Identity

IMS: IP Multimedia Subsystem

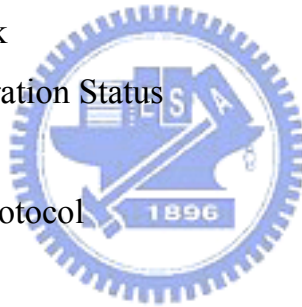
IMSI: International Mobile Subscriber Identity

IN: Intelligent Network

IP: Internet Protocol



ISIM: IMS Subscriber Identity Module
IVR: Interactive Voice Response
MAP: Mobile Application Part
MGC: Media Gateway Controller
MGCP: Media Gateway Control Protocol
MS: Mobile Station
OSA: Open Service Access
P2P: Peer-to-Peer
PBX: Private Branch Exchange
PCM: Pulse-Code Modulation
PS: Packet-Switched
PSTN: Public Switched Telephone Network
PTN: Private Telecommunications Network
RAN: Radio Access Network
RAS: Registration Administration Status
RGW: Residential Gateway
RTP: Real-time Transport Protocol
SCP: Service Control Point
SDP: Session Description Protocol
SER: SIP Express Router
SGSN: Serving GPRS Support Node
SIP: Session Initiation Protocol
SS7: Signaling System Number 7
SSP: Service Switching Point
UE: User Equipment
UMTS: Universal Mobile Telecommunications System
USIM: Universal Subscriber Identity Module
UTRAN: UMTS Terrestrial Radio Access Network
VoIP: Voice over IP



Chapter 1

Introduction

In recent years, there are many forces urging telecom operators to provide the converged services by combining Voice over IP (VoIP) services with advanced data and information services. First, customers hope to enjoy unified services on different type of devices, such as cell phones, personal digital assistants (PDA), smart phones, laptop, or desktop PCs. Second, users want to have unified services regardless of the kind of underlying access network being used, such as WiFi, 3G, Ethernet, WiMax, ADSL, or cable. Third, fixed network providers' profit is reducing, so they must begin to look for the opportunities in wireless domain. Furthermore, mobile operators' speed of growth slows down, and they begin to explore opportunities to expand their markets. Finally, the competition between the fixed and mobile carriers will drive the network service toward convergence, and some successful cases of Internet telecommunications services, such as Yahoo! Messenger, Skype, etc., make Internet Protocol (IP) applications take an important role during the process toward convergence.

The Third Generation (3G) and Beyond the 3G (B3G) mobile systems are developed in order to offer Internet access to mobile users. The Universal Mobile Telecommunications System (UMTS) [25, 27] standardized via the 3GPP represents an evolution in terms of capacity, data speed and new service capabilities from second

generation mobile networks such as Global System for Mobile Communications (GSM) and General Packet Radio Service (GPRS) [25]. Today, more than 122 3G/UMTS networks using WCDMA technology are operating commercially in 55 countries [52].

The Third Generation Partnership Project (3GPP) [53] IP Multimedia Subsystem (IMS) [21] is a standardized next generation networking architecture for telecom operators that want to provide multimedia services over mobile and fixed networks. It supports VoIP communications based on a 3GPP standardized implementation of Session Initiation Protocol (SIP) [6], and runs over the standard Internet Protocol (IP). Although the IMS was originally specified for 3G mobile networks, it can also be provided on any IP-based networks, such as WiFi, corporate enterprise LANs, and the public Internet.

1.1 IP Multimedia Subsystem

The 3GPP proposed the UMTS all-IP architecture to integrate IP and mobile technologies [28]. This architecture evolved from GPRS, UMTS Release 1999 (UMTS R99), and UMTS Release 2000 (UMTS R00). UMTS Release 2000 has been split up into Release 4 and 5. Release 4 introduces a next-generation network architecture for the circuit-switched (CS) domain. Release 5 introduces the IP Multimedia Subsystem (IMS) [21] on top of the packet-switched (PS) domain. The deployment of all-IP network has the following advantages. First, mobile operators could benefit from all existing Internet applications, as well as from the introduction of new services. Second, this evolution allows telecommunications operators to use a common platform (for example, IP) to provide CS and PS domain services, and reduce deployment and operating costs. Third, the new generation of applications will be developed in an all-IP environment, which shortens the distance between mobile and the Internet.

As consumers become increasingly mobile, they will demand wireless Internet access from everywhere. In keeping with these requirements of end users, IMS standards based on the 3GPP UMTS become more and more important. IMS uses Session Initiation Protocol (SIP) [6] defined by IETF and adds extensions for the requirements of mobility to offer multimedia session negotiation and session management over IP. IMS provides an architecture that is independent of the underlying access network, such independency is very important for convergence. Now almost every access network is being enabled to work with an IMS core, including DSL, cable, WiFi, GPRS, WCDMA, or any emerging technology, such as WiMAX. The 3GPP2 group [54] also constructs their Multimedia Domain (MMD) solution based on the IMS network. This will allow CMDA2000 based access network to provide more advanced third generation mobile services.

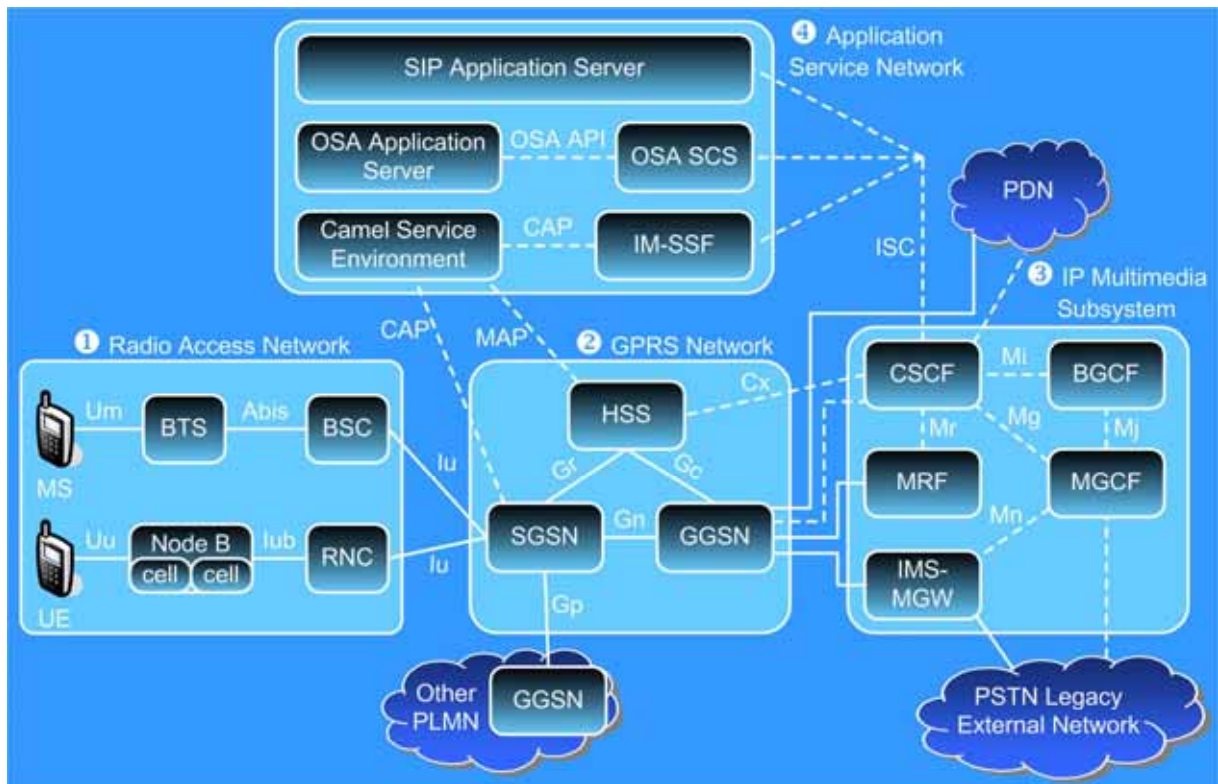
IMS not only provides more innovative services for any type of media session (e.g. voice, video, text, etc.), it also provides the functionality to simultaneously combine CS and PS domain services, and allows sessions to be dynamically modified “on the fly” (e.g. adding a video component to an existing voice session). These capabilities enable IMS to provide a number of new user-to-user and multi-user applications such as enhanced voice services, video telephony, chat, Push-to-talk over Cellular (PoC) [55] and multimedia conferencing.

The UMTS all-IP network architecture [56] consists of the following segments:

Radio Access Network (RAN; see Figure 1.1(1)) can be UMTS Terrestrial Radio Access Network (UTRAN) or GSM Enhanced Data Rates for Global Evolution (EDGE) Radio Access Network (GERAN).

GPRS Core Network (Figure 1.1 (2)) consists of Serving GPRS Support Nodes (SGSNs) and Gateway GPRS Support Nodes (GGSNs) that provide mobility management and session management. Home Subscriber Server (HSS) is the master

database containing all 3G user subscription data. The HSS consists of the IMS functionality (IMS user database) and the Home Location Register (HLR) functionality required by the PS domain and the CS domain to provide support for call handling entities. The Iu interface between UTRAN and SGSN is IP based. SGSN and GGSN communicate with HSS through Gr and Gc interfaces, respectively. These two interfaces are based on Mobile Application Part (MAP).



BGCF: Breakout Gateway Control Function
 BSC: Base Station Controller
 BTS: Base Transceiver Station
 CAMEL: Customized Application
 Mobile Enhanced Logic
 CAP: CAMEL Application Part
 CSCF: Call Session Control Function
 GGSN: Gateway GPRS Support Node
 HSS: Home Subscriber Server
 ISC: IMS Service Control
 MAP: Mobile Application Part

MGCF: Media Gateway Control Function
 MGW: Media Gateway
 MRF: Media Resource Function
 MS: Mobile Station
 OSA: Open Service Access
 RNC: Radio Network Controller
 SCS: OSA Service Capability Server
 SGSN: Serving GPRS Support Node
 SSF: Service Switching Function
 T-SGW: Transport Signaling Gateway
 UE: User Equipment

Figure 1.1. UMTS All-IP Network Architecture.

IP Multimedia Subsystem (IMS; Figure 1.1 (3)) is located between the GGSN and the PDN (specifically, the IP networks). In this subsystem, the Call Session Control

Function (CSCF) [21] is a SIP server, which is responsible for call control. Other nodes in the IMS include Breakout Gateway Control Function (BGCF), Media Gateway Control Function (MGCF), and IM-Media Gateway Function (IM-MGW). These nodes are typically used in a VoIP network.

Application and Service Network (Figure 1.1 (4)) supports flexible services through a service platform. The IMS network architecture will provide a separation of service control from call/connection control, and the applications are implemented in dedicated application servers that host service-related databases or libraries. The IMS Service Control (ISC) interface is the IETF SIP protocol with extensions for service control. The 3GPP defines three possible alternatives to provide flexible and global services:

- SIP Application Server: The SIP application services are either developed by the mobile operators or from trusted third parties.
- IM-Service Switching Function (IM-SSF) and CAMEL Service Environment (CSE): This server will be used by the mobile operator to provide CAMEL services to the IMS users. The legacy CS domain Services are provided via the CSE platform.
- Open Service Access (OSA) Service Capability Server (SCS) and OSA Application Server: This server will be used to give third parties controlled access to the operator's network, and enable third parties to run their own applications (in the third-party application servers) using the IMS capabilities of the operator's network.

In the IMS network, a CSCF can be proxy, serving, or interrogating. The Proxy CSCF (P-CSCF) is the first contact point within the IMS for a User Equipment (UE). The P-CSCF may be located in the home or visited network. The P-CSCF ensures that registration of the user is passed to the correct home network and that SIP session

messages are passed to the correct Serving CSCF (S-CSCF) once registration has occurred. The P-CSCF is an important function as it is in the position to detect services. The S-CSCF is the function that registers the user and provides service to them. It performs routing and translation, provides billing information to mediation systems, maintains session timers, and interrogates the HSS to retrieve authorization, service triggering information and user profile. In short, it is the brain of the IMS. Interrogating CSCF (I-CSCF) is the function that is able to determine the S-CSCF with which a user should register. This is achieved by querying the HSS, which checks that the user is allowed to register in the originating network and returns an S-CSCF name and capability if this is the case. The I-CSCF is then able to contact the S-CSCF with the register. The I-CSCF function can be removed from the signaling path once it has been used to establish which S-CSCF is in use. The exception to this is if the THIG (Topology Hiding Inter-network Gateway) function of the I-CSCF is being used.

In this dissertation, we study the UMTS all-IP network in three parts. First, an integrated Call Agent of the converged VoIP network is presented. This topic focuses on the IMS network components (Figure 1.1 (3)) and discusses the interworking between different VoIP protocols, such as SIP, MGCP and H.323. After the IMS network and the VoIP protocols interworking are elaborated, we study the correlation between IMS and GPRS networks. We find the authentication procedures of GPRS and IMS networks are almost the same. To eliminate the redundancy of the authentication procedures, we proposed the one-pass GPRS and IMS authentication procedure for UMTS. This topic discusses the authentication procedure started from the UE of the radio access network (Figure 1.1 (1)) through the GPRS core network (Figure 1.1 (2)) to the IMS network. After the GPRS and IMS networks, we study the application platform of the 3GPP all-IP network architecture. A SIP based call center with waiting time prediction is elaborated. This work focuses on the service platform

(Figure 1.1 (4)) and uses the call center as a SIP application server.

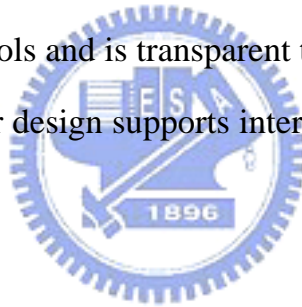
1.2 An Integrated Call Agent of the Converged VoIP Network

The traditional circuit-switched telecommunication network and the packet-switch data network are converging to a single packet-switched multimedia service network. One important application of the convergence is IP telephony communications, also referred to as VoIP (Voice over IP). In a converged VoIP network such as IMS, media gateways are responsible for connecting packet-switched network with the PSTN network. Media gateway controllers (MGCs) also referred to as Call Agents (CA) perform call control functions and instruct commands to media gateways through the Media Gateway Control Protocol (MGCP) [7]. Other VoIP protocols including SIP and H.323 provides multimedia call control over packet-switched network. H.323 is an ITU-T recommendation for multimedia conferencing [4]. It is a protocol umbrella that consists of many standards, including Q.931 call control protocol, H.225 registration and administration, and H.245 media negotiation. SIP has been standardized by IETF for interactive communication sessions between users. It as well is chosen as the IMS Service Control (ISC) interface between the CSCF and the service platform in the UMTS IMS network. SIP is a text-based request-response architecture similar to HTTP. Since SIP is simpler and more flexible then H.323, many consider SIP a powerful alternative to H.323. H.323 and SIP utilize Real-time Transport Protocol (RTP) [36] to transmit real-time packets over the Internet.

To provide interworking functions in the converged VoIP network and enable devices using different protocols to communicate, gateways are needed to translate

messages between different protocols. Vemuri described an inter-operation model called SPHINX (SIP, H.323 and IN interworking) [11] where H.323 and SIP terminals can access IN services. In addition, an SIP-H323 gateway is a byproduct of this inter-operation model based on the half-call state model of the IN. Gurbani and Rastogi [14] and Haerens [15] suggested ways to map the call control of SIP to IN, but they did not support the H.323 slow-start call setup. Agrawal, Schulzrinne and Singh specified the requirements for SIP-H.323 interworking [12, 13]. However, no work has been done on interoperating all VoIP protocols in a simple and flexible framework.

In this topic, we present an integrated call agent architecture that supports the interworking function of VoIP protocols (SIP, H.323, MGCP and MEGACO) using the basic call state model in Intelligent Network. The interworking function translates messages of the VoIP protocols and is transparent to the call parties and kept as simple as possible. Furthermore, our design supports inter-CA communications using SIP in a straightforward manner.



1.3 One-Pass GPRS and IMS Authentication

Procedure for UMTS

The packet data services of the UMTS PS core network are provided by the SGSN via UTRAN. The SGSN connects to the external data network through the GGSN. Furthermore, the SGSN communicates with the home subscriber server (HSS) and the authentication center (AuC) [56] to retrieve subscriber data and authentication information of an MS. When an MS performs location update or tries to attach onto the GPRS service, the SGSN may perform GPRS authentication to the user. The authenticating parties are HSS in the home network and the universal subscriber identity module (USIM) in the MS. GPRS authentication consists of two major

procedures:

- **Distribution of authentication information:** The SGSN retrieves an array of authentication vectors (AVs) of the MS from the AuC.
- **Authentication and key agreement (AKA):** This procedure uses the AV which is obtained from the previous step to perform mutual authentication between an MS and the network by showing knowledge of a pre-shared secret key that is only available in the USIM of the MS and in the AuC.

The detail of the GPRS authentication will be elaborated in chapter 3. In addition to GPRS authentication, it is necessary to authenticate the MS before it can access IMS services. Without IMS authentication, a mobile user who passes the GPRS authentication can easily fake being another IMS user. The IMS authentication procedure is performed by the IMS registration which is invoked from the MS through the P-CSCF to the S-CSCF. Since the IMS information is delivered through the GPRS transport network, a UMTS mobile station must activate GPRS packet data protocol (PDP) context [20] before it can register to the IMS network. The 3GPP specifications define mutual authentication mechanisms [19] in both the GPRS network and the IMS. Although both GPRS and IMS authentications are necessary, most steps in this 3GPP “two-pass” authentication procedure are duplicated. G. Horn, D. Kröselberg, and K. Müller [57] surveyed the security architecture of the IMS including user specific features protecting the access of the IMS user, such as AKA when a user registers, integrity protection of IMS access signaling and independent protection of SIP signaling in the IMS core network. Y.-B. Lin and Y.-K. Chen [23] surveyed the authentication procedure of UMTS and proposed an algorithm to reduce the authentication signaling traffic. However, no work has been done on the reduction of the duplicated two-pass authentication procedure of GPRS and IMS.

Based on our observation, we propose a one-pass authentication procedure that only needs to perform GPRS authentication. At the IMS level, authentication is implicitly performed in IMS registration. Our approach may save up to 50% of the IMS registration/authentication traffic, as compared with the 3GPP two-pass procedure. We formally prove that the one-pass procedure correctly authenticates the IMS users.

1.4 A SIP-based Call Center with Waiting Time

Prediction

A call center is a centralized office that answers incoming telephone calls from customers or makes outgoing telephone calls to customers. When all agents are busy, the arriving customer calls have to wait in a queue. A waiting customer will be served after an agent becomes available. To improve customer satisfaction, it is essential to inform the waiting customers of predicted delays before they can be served. Whitt [43, 44] proposed an algorithm to predict the waiting time for customers. It is important to evaluate the accuracy of prediction, which was not mentioned in [43].

In this topic, we propose a plug-in modular architecture for an Internet call center with waiting time prediction. We implement this plug-in module in the SIP Express Router (SER) [40] platform. The SER is a high-performance, configurable, free SIP server. It can act as a SIP registrar, proxy or redirect server. The plug-in SER module called Automatic Call Distributor (ACD) module consists of five components: the agent dispatcher, the queue manager, the agent state manager, the Interactive Voice Response (IVR) and the ACD database.

The ACD module maintains a status record for each of the agents. The record is accessed by the agent state manager. A five-state Finite State Machine (FSM) is

associated with the record. We describe the detailed functionality and the interfaces of each component in Chapter 4.

Then we propose two output measures and develop a discrete event simulation model to investigate the performance of the waiting time prediction algorithm for the call center. Our study indicates that the waiting times can be more accurately predicted when the call arrival rate is large.

This dissertation is organized as follows. Chapter 2 presents an integrated Call Agent of the converged VoIP network. Chapter 3 describes the one-pass GPRS and IMS authentication procedure for UMTS. In Chapter 4, we describe a SIP based call center with waiting time prediction. Chapter 5 concludes this dissertation and describes the future work.



Chapter 2

An Integrated Call Agent of the Converged VoIP Network

The traditional circuit-switched telecommunication network and the packet-switched data network are converging to a single packet-switched network. One important application of the converged network is telephony communications, also referred to as VoIP (Voice over IP). Call signaling protocols, such as H.323, SIP and MGCP, have been developed to support VoIP communications. To enable devices using different VoIP protocols to communicate, gateways are needed to translate messages of one protocol to messages of another. In this chapter, we present a simple, flexible framework for this interworking function. The framework is based on a half-call model where a call is controlled by two half-call finite state machines (FSMs), one representing the state of the caller and the other representing the state of the callee. The interworking function has been implemented such that the caller FSM of one VoIP protocol can interact with the callee FSM of any VoIP protocol. The development effort of the interworking function is minimized since only two half-call FSMs for each VoIP protocol are needed and they can be developed independently as long as the design conforms to the same interface specification. We have developed an integrated call agent

(ICA) that contains the half-call FSMs of H.323, SIP and MGCP. Calls between devices using these VoIP protocols can be set up, maintained and terminated by the ICAs.

2.1 Introduction

The PSTN (Public Switched Telephone Network) has provided reliable voice communication for decades. A telephone call to anyone in the world can be established in seconds, and the voice quality is good in general. Voice waveform transmitted in the PSTN is encoded using PCM (pulse-code modulation, G.711 A-law and u-law, both 64kbps) technique. To establish a telephone call between two parties, a dedicated link needs to be set up, and the link has to be torn down when the call terminates. This work is performed by telephone switches exchanging standard signaling, such as ISUP (Integrated Services Digital Network User Part).

2.1.1 VoIP protocols

As the Internet becomes overwhelmingly widespread, transporting voice communication traffic using the Internet Protocol (IP) provides advantages over the traditional PSTN. This is often referred to as IP telephony or VoIP (Voice over IP). VoIP can use sophisticated speech codecs, such as G.723.1 and G.729 [1], to reduce the bandwidth required for a call. VoIP communications use RTP/RTCP to transport voice packets over IP networks [2,3]. To establish a call, the two parties

involved should negotiate the codec and the RTP/RTCP ports used for the call, as well as exchange messages to set up and terminate the call. H.323 and SIP are two existing VoIP signaling protocols. Both are designed to support the setup, communication capacity exchange and tear-down of a VoIP call.

H.323 is an ITU-T recommendation for multimedia conferencing over packet-switched networks [4]. It is a protocol umbrella that consists of many standards, including Q.931 call control protocol, H.225 registration and administration, and H.245 media negotiation. It is widely used nowadays; NetMeeting of Microsoft and VoIP products from Cisco all support H.323 multimedia communications over packet-switched networks. H.323 also defines a gatekeeper as a manager and arbiter over a network. The gatekeeper is an optional entity in charge of endpoint registration, address translation, and bandwidth assignment. An H.323 endpoint can make a direct or gatekeeper-routed call. One of the major criticisms against H.323 is the time and complexity involved in setting up a call; H.323 version 1 uses multiple stages to exchange signaling and media capabilities. Moreover, messages are transported using TCP, which requires additional session set-up time. Recent H.323 versions can include both signaling and media capabilities in a single message transmitted by UDP, which eliminates the additional round-trip time for TCP handshake [5]. However, with too many options, H.323 still has compatibility problems for the products from different providers.

Session Initiation Protocol (SIP) has been standardized by IETF for initiating interactive communication sessions between users [6]. It can be used to establish Internet telephony calls. SIP is a lightweight protocol, which uses only six commands for session control. With early media capability exchange feature, a

SIP terminal can issue only one command, INVITE, to set up a call. Therefore, it is faster than H.323 in call establishing. SIP is a text-based request-response architecture similar to HTTP. Since SIP is simpler than H.323, many consider SIP a powerful alternative to H.323

To enable communications between VoIP users and PSTN users, gateways between the PSTN and IP-based networks are needed. Since the signaling used in the VoIP networks and the PSTN are different, and the voice media are transmitted in different formats, the gateways have to perform two functions: signaling conversion and media conversion. The two functions can be carried out in two separated entities: MGC (Media Gateway Controller) and MG (Media Gateway). An MGC (also referred to as a CA, Call Agent) performs the signaling conversion function, and an MG performs the media conversion function. A standardized protocol can be used between an MGC and MGs. IETF proposed the MGCP (Media Gateway Control Protocol) architecture depicted in Figure 1 [7]. MGCP is a master-slave protocol. A CA is a master and has control over several MGs; an MG acts as a slave and is kept simple and passive. A CA also performs the call control function as a gatekeeper in H.323, but has much tighter control. A CA instructs an MG to establish, maintain, and terminate a call between a VoIP terminal and an endpoint in the PSTN. MGCP is based on a centralized network infrastructure. To serve to a wide area network, a group of CAs need to coordinate, but MGCP does not specify how the CAs interact. Signaling used in inter-CA communications can be SIP or ISUP. MGCP describes the media capability and parameter using SDP (Session Description Protocol, [8]). Recently Megaco (or H.248), an enhanced version of MGCP, is promoted jointly by the IETF and ITU-T [9].

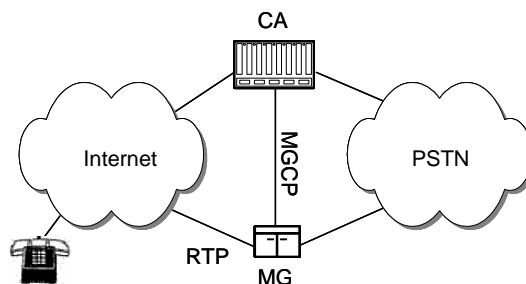


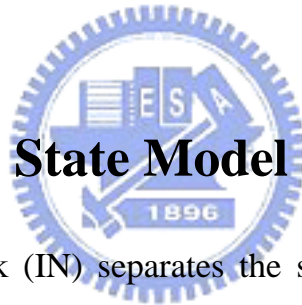
Figure 1. MGCP architecture.

2.1.2 Interoperation

Since H.323 and SIP are expected to co-exist in the near future, gateways between H.323 and SIP terminals are needed for the interworking function between H.323 and SIP. EURESCOM proposed a project for providing IN functionality for H.323 telephony calls [10]. Vemuri described an inter-operation model called SPHINX (SIP, H.323 and IN interworking) where H.323 and SIP terminals can access IN services [11]. In addition, an SIP-H323 gateway is a byproduct of this inter-operation model based on the half-call state model of the IN. Agrawal, Schulzrinne and Singh [12,13] specified the requirements for SIP-H.323 interworking. Gurbani and Rastogi [14] and Haerens [15] suggested ways to map the call control of SIP to IN, but they did not support the H.323 slow-start call setup. Ackermann, et al., have implemented a sip-h323 gateway as a basis for supplementary service interworking [16]. They use a dedicated call model that transfers H.323 messages to SIP messages, and vice versa. Singh and Schulzrinne have presented a Columbia InterNet Extensible Multimedia Architecture (CINEMA) where MGCP and H.323 can be interworking through SIP [17]. Nevertheless, it is difficult to modify this call model to cooperate with other VoIP protocols.

However, no work has been done on interoperating all VoIP protocols in a simple and flexible framework. In this chapter we present an integrated call agent architecture that supports the interworking function of VoIP protocols (SIP, H.323, MGCP and MEGACO) using the basic call state model in Intelligent Network [18]. The interworking function translates messages of the VOIP protocols. The translation is transparent to the call parties and kept as simple as possible. Furthermore, our design supports inter-CA communications using SIP in a straightforward manner. The rest of the chapter is organized as follows. Section 2.2 briefly describes the IN basic call state model. Section 2.3 presents the system design and the implementation issues. Implementation issues and results are discussed in Section 2.4. Conclusions are given in Section 2.5.

2.2 IN Basic Call State Model



An intelligent network (IN) separates the service logic from the switching function in the telecommunications network; the service intelligence is placed in computer nodes that are distributed throughout the network. This provides the network operators with the means to develop and control services more efficiently. New capabilities can be rapidly introduced and customized for the network. A simple IN architecture is depicted in Figure 2. A service switching point (SSP) is an IN-capable switching system dealing with the call control functions that establish, maintain, and clear a call. In addition, it detects user requests for IN-based services and queries a service control point (SCP) to determine how the call should be handled. The SCP contains the service logic to provide the IN services.

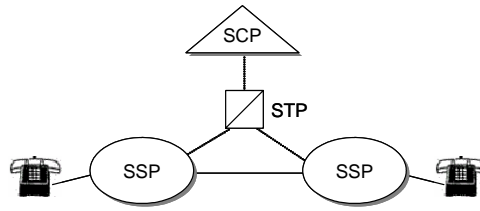


Figure 2. IN architecture.

The basic call control function of an SSP is supported by a finite state machine (FSM) called basic call state model (BCSM). The BCSM consists of point in calls (PICs), detection points (DPs), and events. PICs represent the switching activities or states that a call goes through from call origination to termination. DPs are states at which transfer of control from the SSP to the SCP can occur. Events are messages exchanged between BCSMs, and trigger state transitions of the BCSMs.

Figure 3 shows that the BCSM is based on half-call model; a call model consists of two half-call models: the originating BCSM (O_BCSM) and terminating BCSM (T_BCSM). The O_BCSM represents the states associated with the call originating party; the T_BCSM represents the states associated with the call terminating party. Note that the originating and terminating BCSMs interact with the call parties using ISUP messages. The call control functions are performed through the exchange of events between the O_BCSMs and T_BCSMs. These events include Setup, Alert, Answer, Disconnect, Busy, No-Answer, and Abandon. For details, the readers are referred to [18].

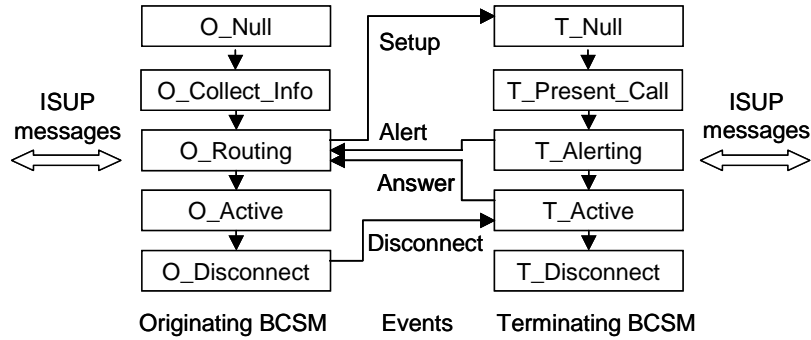


Figure 3. Simplified IN BCSMs.

2.3 System Architecture

We proposed a simple systematic way to implement the interworking function (IWF) between VoIP protocols. There are two major functions in the IWF: call routing function and call control function. The call routing function locates the called party by querying a location database so that the call request can be delivered. How to locate a user is beyond the scope of this paper; we assume that a user location database exists. The call control function sets up and maintains the call by translating messages between two VoIP protocols.

The call control function can be implemented using the BCSMs. For each VoIP protocols, its messages are mapped to the messages of the BCSMs. An interworking gateway of two VoIP protocols can be implemented by combining the BCSMs of the VoIP protocols. Therefore, an O_BCSM of one VoIP protocol and a T_BCSM of another VoIP protocol can interact through the exchange of a set of unified events based on the IN half-call model. Figure 4 shows the six BCSMs of three major VoIP protocols for a general gateway design. Note that the events between an O_BCSM and a T_BCSM are protocol independent. This design reduces developing efforts. To interwork n different VoIP protocols, each

protocol needs only one O_BCSM and one T_BCSM (i.e. $2n$ BCSMs in total), instead of hardcoding n^2 messages translation modules between any two protocols.

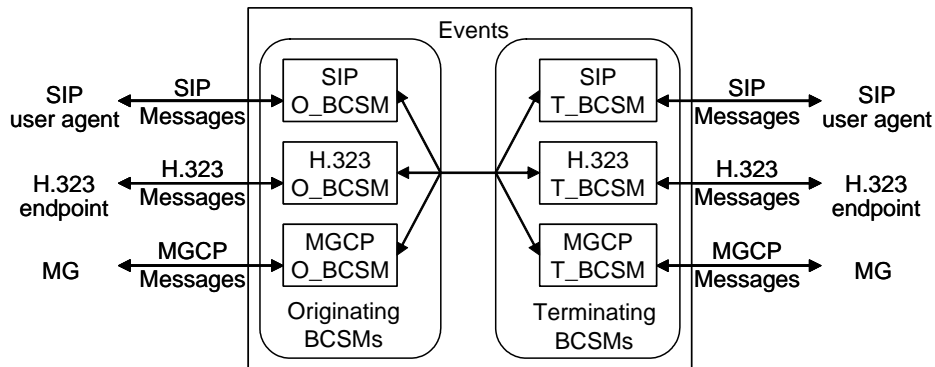


Figure 4. Components developed for a general VoIP gateway.

A combination of the BCSMs of two VoIP protocols becomes a gateway; the construction of a gateway is to design the BCSMs of the VoIP protocols. Figure 5 shows an example SIP/H.323 gateway. This gateway consists of the BCSMs of SIP and H.323. When a SIP UA originates a call to a H.323 terminal, a SIP O_BCSM receives a SIP INVITE message and issues a Setup event to an H.323 T_BCSM. After being notified by the Setup event, the H.323 T_BCSM sends an H.323 SETUP message to the terminating H.323 endpoint. Subsequent SIP and H.323 messages are exchanged to set up the call through the interaction of the SIP O_BCSM and H.323 T_BCSM. On the other hand, an originating call from a H.323 terminal to a SIP UA can be handled by an H.323 O_BCSM and a SIP T_BCSM.

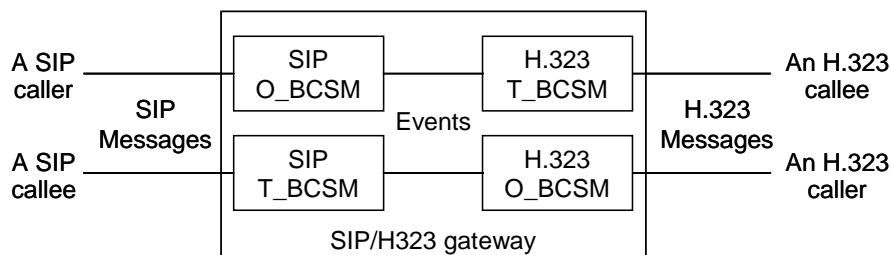


Figure 5. A SIP/H.323 gateway.

All the gateways needed in the converged VoIP network can be implemented

in the same way described above. Figure 6 depicts a converged network using three gateways: SIP/H.323, SIP/MGCP, and MGCP/H.323 gateways. Note that the gateways can share the same BCSMs developed. Moreover, when a new VoIP protocol, such as MEGACO, is added to the converged network, the gateways to MEGACO network use the same BCSMs developed for existing VoIP protocols; only the BCSMs of MEGACO need to be implemented.

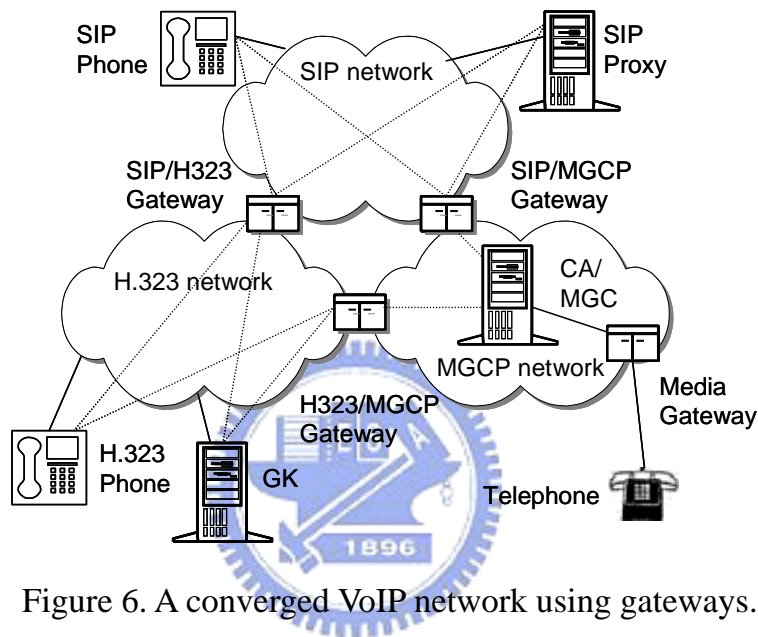


Figure 6. A converged VoIP network using gateways.

To route an originating call, the location of the callee and the signaling protocol that the callee supports must be determined so that a correct T_BCSM can be invoked for messages translation. Solutions to this call routing function are beyond the scope of this paper; we assume that a solution is available. We integrate the call routing function and the BCSMs of all VoIP protocols in an entity called integrated call agent (ICA). The ICA can not only translate messages between VoIP protocols but also act as a H.323 gatekeeper, a SIP proxy/registrar, and a MGCP call agent.

The design is extensible. Figure 7 shows the converged architecture partitioned into two zones managed by two ICAs. Calls between entities of two

zones can be set up by the ICAs exchanging SIP messages through two SIP BCSMs. Figure 8 depicts an example call between an H.323 terminal and an MGCP phone, and two ICAs (ICA 1 and ICA 2) are involved. When an H.323 terminal in Zone 1 initiates a call to an MGCP phone in Zone 2, ICA 1 invokes a SIP T_BCSM and sends a SIP INVITE message to ICA 2. ICA 2 initializes a SIP O_BCSM to handle this call request. Since ICA 2 knows that the callee is an MGCP phone, an MGCP T_BCSM was invoked to set up this call connection.

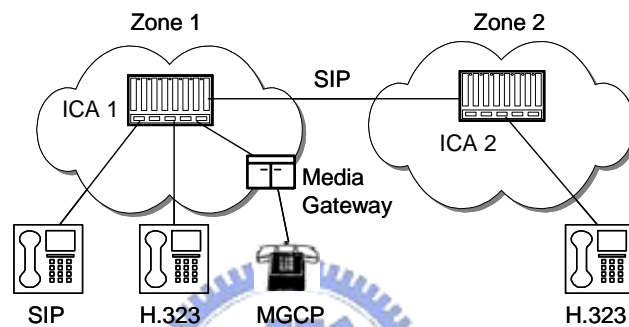


Figure 7. A converged VoIP network managed by integrated call agents.

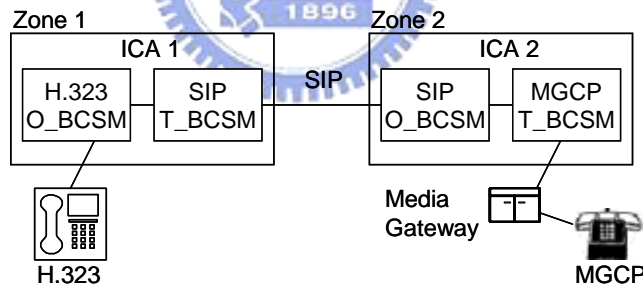


Figure 8. An example of H.323 and MGCP interworking using 2 ICAs.

2.3.1 Mapping of VoIP protocol messages to the BCSM

messages

The implementation of half-call BCSM for each VoIP protocol is to map the VoIP call signaling messages to the IN BCSM messages. We focus on the relationship between the BCSM states and the VoIP messages. This message

mapping is straightforward except for the slow-start version of H.323. Figure 9 depicts the mapping for H.323, SIP, and MGCP.

When a call request message arrives (e.g., **INVITE** for SIP, **SETUP** for H.323, or **NOTIFY** for MGCP) from a calling party, the corresponding O_BCSM becomes active and initialize itself to state *O_Null*. The O_BCSM proceeds to state *O_Collect_Info* where a called number or address is collected from the calling party. For SIP and H.323, the **INVITE** and **SETUP** messages carry the identifier of the called party, media descriptor, and signaling transport address. Hence the O_BCSM simply skips state *O_Collect_Info* and go to state *O_Routing* directly. For MGCP, however, a **RQNT** message should be sent to the calling party to collect the dialed number. Once the calling party collects the number dialed, it sends a **NTFY** with the dialed number to the O_BCSM. Then, the O_BCSM transits to state *O_Routing*.

State *O_Routing* responds to the calling party with the call progressing message (e.g., **100 Trying** for SIP or **CALLPROC** for H.323). In addition, the VoIP protocol used by the called party is determined, and a **Setup** event is issued. Consequently, a T_BCSM for the called party is initialized at state *T_Null* to handle this **Setup** event. The O_BCSM stays in state *O_Routing* waiting for the **Alert** and **Answer** events sent from the T_BCSM. If both events are received, the call has been accepted by the called party and the O_BCSM responds with a call connect message (such as **CONNECT** for H.323, **OK** for SIP, or **MDCX** for MGCP) to the calling party indicating that the call has been accepted. The O_BCSM stays at state *O_Active* until the call is terminated.

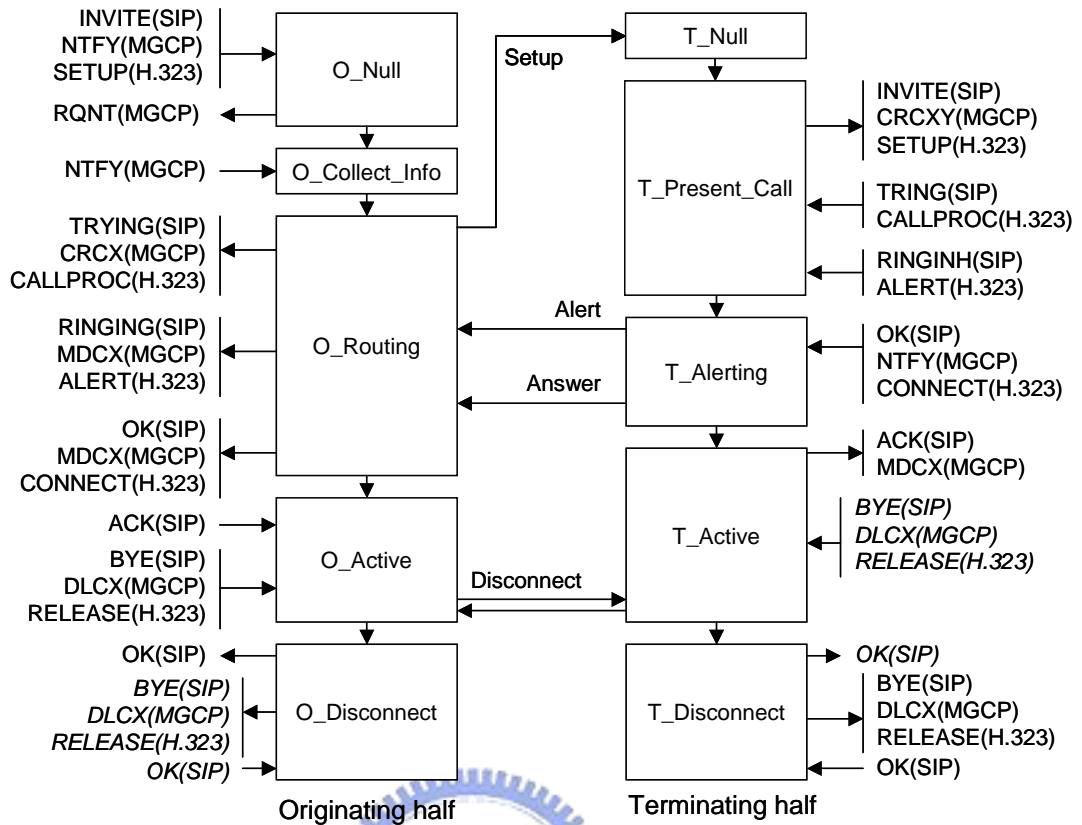


Figure 9. Mapping VoIP messages to BCSM messages.

When a T_BCSM is activated by a Setup event from an O_BCSM, it goes to state *T_Present_Call* and checks the validation of the called number. If the number is valid, the T_BCSM sends a call request message (e.g., INVITE for SIP, SETUP for H.323, or CRCX for MGCP) to the called party and waits for the response messages such as trying, altering, and answered from the called party. Upon receiving the alerting message (e.g., 180 Ringing for SIP or ALERT for H.323), the T_BCSM informs the O_BCSM with an Alert event and proceeds to state *T_Alert*.

In state *T_Alert*, the T_BCSM waits for the call connect message (e.g., OK for SIP, CONNECT for H.323, or NTFY for MGCP). Once the message is received, the call has been accepted by the called party. The T_BCSM activates an Answer event to the O_BCSM and transits to state *T_Active*. In this state, an

acknowledgement message (e.g. ACK for SIP or MDCX for MGCP) is sent back to the called party. Thus far, a basic two party call connection is established.

If a disconnect message (e.g., BYE for SIP, RELEASE for H.323, or DLCX for MGCP) is received by either the originating or the terminating BCSM, the BCSM will issue a *Disconnect* event to the corresponding BCSM of this call. Both BCSMs proceeds to state *Disconnect*, and the call is terminated.

2.3.2 H.323 slow-start

For the H.323 slow-start, the mapping is quite different and those events described above are inadequate to exchange the media capability. For the BCSMs described above, after the media capability of the calling party becomes available at state *O_Null*, the O_BCSM issues a *Setup* event. Moreover, after the capability of the called party becomes available at state *T_Present_Call*, the T_BCSM issues an *Answer* event respectively. However, for H.323 slow-start, this capabilities will not be determined until the H.245 negotiation after the *CONNECT* messages. As a result, the BCSMs do not get media capability when the *Setup* or *Answer* event is issued. In addition, there is no event reporting the media capability to the calling and called party after the *Answer* event. Thus, we made two modifications to the H.323 BCSMs. First, the *Answer* event is postponed to state *T_Alert* where the T_BCSM receives an H.245 open logical channel message (OLC) which carries the media capability for the called party. Second, a new event, media capability ready (*MediaReady*), is introduced at the end of state *O_Routing* to notify the T_BCSM the media capability of the calling party. Figure 10 shows the modifications to the mapping for the H.323 slow-start.

To interwork with the H.323 BCSMs that supports slow-start feature, the

BCSMs of SIP and MGCP were modified to send a **MediaReady** event anyway at the end of state *O_Routing*. i.e., all O_BCSMs should issue a **MediaReady** event at state *O_Routing*. Although the called party of SIP or MGCP expects to receive the media capability when it receives the **INVITE** or the **CRCX** message, this capability exchange can be delayed till when the **ACK** or the **MDCX** message is received. Figure 11 shows the call flow of H323 and SIP interworking where the H.323 terminal is in slow-start mode. The call flow of a slow-start H323 terminal and an MGCP phone can be done in a similar manner. If the calling party is an H.323 phone using the slow-start feature, the T_BCSM issues an **INVITE** message without specifying the caller's media capability to the SIP callee. The media capability is available at state *O_Routing* after the H.245 OLC message is received. Consequently, the O_BCSM issues a **MediaReady** event to the T_BCSM at the state *O_Routing*. Therefore, the T_BCSM encloses this capability in the **ACK** message sent to the called party.

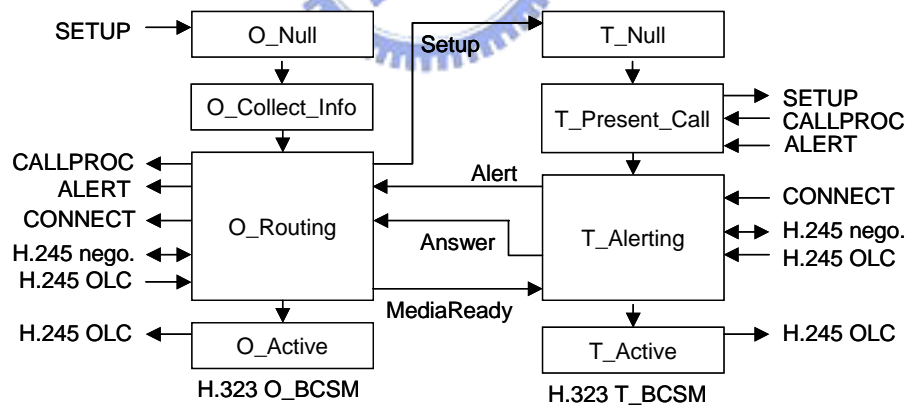


Figure 10. BCSMs for the H.323 slow-start.

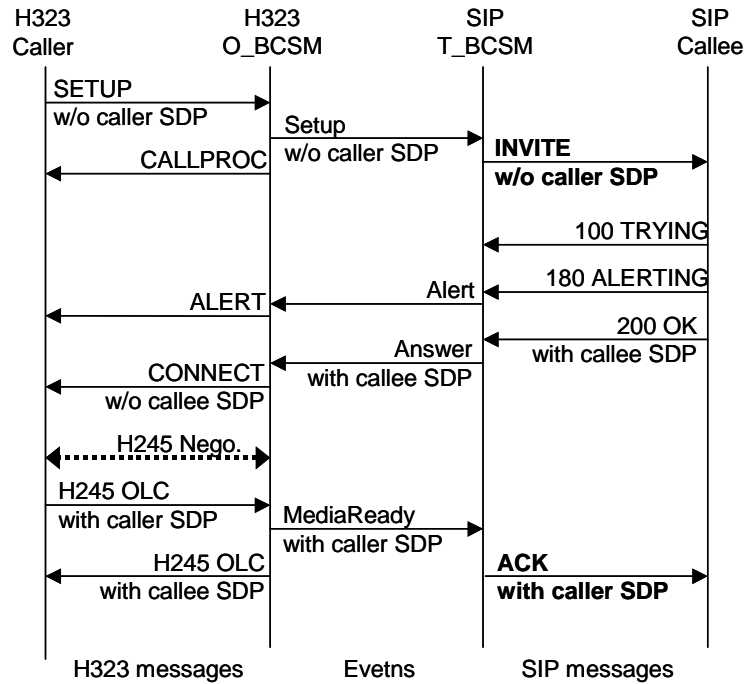


Figure 11. Call flow of H.323 (slow-start) and SIP interworking.

2.4 System Implementation and Result

To reduce the effort in developing a VoIP gateway, we use existing, well-developed protocol stacks and endpoints. In our implementation, we use the MGCP and SIP protocol stacks developed by CCL/ITRI, Taiwan and the open-source H.323 protocol stack developed by OpenH323. In addition, our experiment platform includes two residential gateways (RGWs) which are also developed by ITRI using D/41E and D/41ESC cards from Dialogic Corp. The RGWs support MGCP and can connect up to 16 telephones. The ICA platform also supports Microsoft NetMeeting (using H.323) and SIP user agent. The H.323 BCSMs are modified from the source code of the OpenH323's OpenGate that supports registration administration status (RAS) messages and gatekeeper-routed call signaling. We have also developed a SIP proxy/registrar based on the ITRI

SIP protocol stack. Figure 12 summarizes the components used in our platform.

In our experiments, a call can be successfully set up between any two VoIP phones. The Microsoft NetMeeting currently supports only H.323 slow-start version; we use an open-source OpenPhone (with both slow-start and fast-start capabilities) to test the cases of slow-start version. In addition, a Vocal sip proxy, developed by Vovida, was used to test calls between SIP UAs. Since an ICA acts as both SIP proxy and H.323 gatekeeper, an ICA can initiate a call to the phones that are under the control of a SIP proxy or an H.323 gatekeeper without further modification.

The comparison of the delays in establishing a call between various types of phones by OpenGate H.323 gateway, our ICA, and Vocal SIP proxy is depicted at Figure 13. No result of inter-protocol calls through OpenGate and Vocal is listed, because they do not convert the messages of different protocols. Figure 13.a shows call establishment delays for the calls initiated from various types of phones to NetMeeting, which only equipped with slow-start mode, and Figure 13.b shows those for calls to OpenPhone in H.323 fast-start Mode. Although our ICA supports signaling conversion for different VoIP protocols, the results indicates that the ICA sets up calls faster than the OpenGate in all cases except for calls between two NetMeeting users.

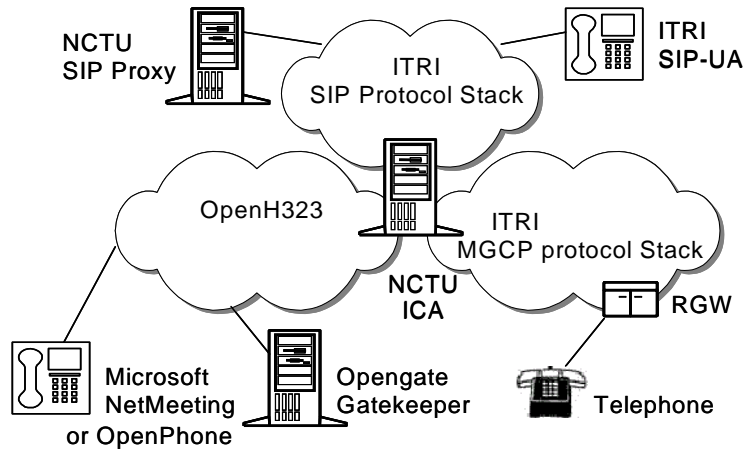


Figure 12. Components used in our platform.

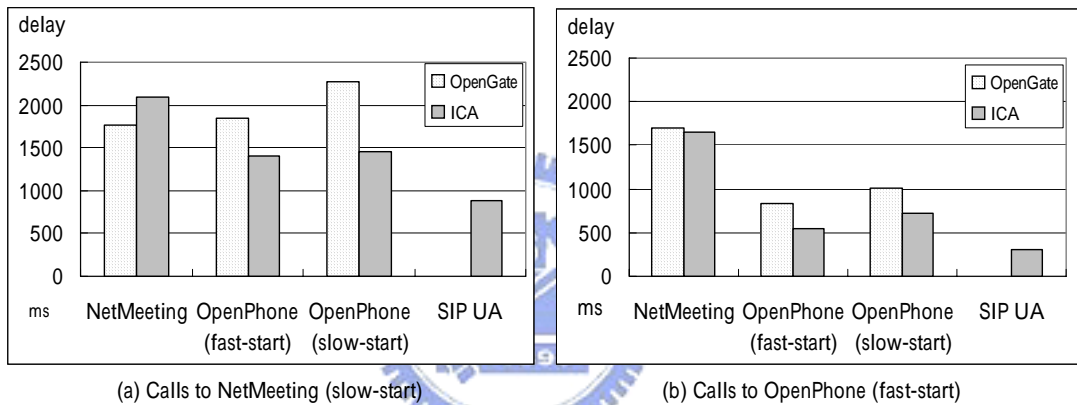
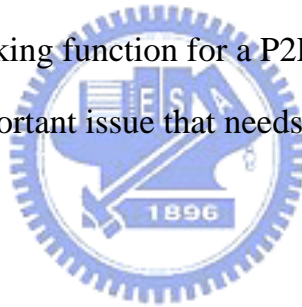


Figure 13. Call establishment delays.

2.5 Conclusions

We have presented a simple, flexible framework for the interworking functions of VoIP protocols based on IN half-call BCSM. In addition, we have implemented the basic gateway components, O_BCSMs and T_BCSMs, for SIP, H.323, and MGCP. Using these components, gateways for SIP/H.323, SIP/MGCP, and MGCP/H.323 can be constructed. This approach simplifies the effort in interworking with a call signaling protocol, such as ISUP and Q.931, in the

network. By using the same interaction events of the half-call model, the BCSMs of a call signaling protocol can interwork with the existing BCSMs. In addition, an ICA containing all the BCSMs is able to translate messages between call signaling protocols. Under this half-call control framework, a converged VoIP network can be managed by a group of coordinating ICAs such that two user devices managed by different ICAs can communicate. The call routing function that determines the location and protocol of the called party has not been fully investigated in this paper. As a mobile user may change his IP address and VoIP devices constantly, this problem becomes even more complicated. We need registration and/or paging schemes to track mobile users in the converged telecommunication network. Recently, P2P (peer-to-peer) VoIP communications, such as Skype, have become very popular. The interworking function for a P2P VoIP system and a client-server one (such as SIP) is an important issue that needs to be investigated.



Chapter 3

One-Pass GPRS and IMS Authentication

Procedure for UMTS

Universal Mobile Telecommunications System (UMTS) supports Internet protocol (IP) multimedia services through IP multimedia core network subsystem (IMS). Since the IMS information is delivered through the general packet radio service (GPRS) transport network, a UMTS mobile station (MS) must activate GPRS packet data protocol (PDP) context before it can register to the IMS network. In the Third-Generation Partnership Project (3GPP) specifications, authentication is performed at both the GPRS and the IMS networks before an MS can access the IMS services. We observe that many steps in this 3GPP “two-pass” authentication procedure are identical. Based on our observation, this chapter proposes a one-pass authentication procedure that only needs to perform GPRS authentication. At the IMS level, authentication is implicitly performed in IMS registration. Our approach may save up to 50% of the IMS registration/authentication traffic, as compared with the 3GPP two-pass procedure. We formally prove that the one-pass procedure correctly authenticate the IMS users.

3.1 INTRODUCTION

Universal Mobile Telecommunications System (UMTS) proposed by the Third-Generation Partnership Project (3GPP) is a third-generation (3G) mobile telecommunications technology evolved from general packet radio service (GPRS) [20]. Fig. 3.1 illustrates the UMTS packet switched (PS) core network (CN), where the packet data services of a mobile station (MS) are provided by the serving GPRS support node (SGSN) via UMTS terrestrial radio access network (UTRAN). The SGSN connects to the external data network through the gateway GPRS support node (GGSN). Furthermore, the SGSN communicates with the home subscriber server (HSS) and the authentication center (AuC) to retrieve subscriber data and authentication information of an MS. The AuC, which may be collocated with the HSS, is responsible for security management of subscribers. UMTS supports voice and multimedia services through the PS CN based on the Internet Protocol (IP) technology. Specifically, the 3GPP defines IP multimedia core network subsystem (IMS) to support multimedia services such as voice telephony, video, real-time interactive games, messaging, and multimedia conferencing [21]. In IMS, multimedia services are provided by call session control function (CSCF) utilizing session initiation protocol (SIP) [6],[22]. Three types of CSCFs are defined in IMS: A proxy-CSCF (P-CSCF) located in the visited network of an MS is responsible for redirecting the SIP messages of an MS to the home network (where the HSS/AuC resides). A serving-CSCF (S-CSCF) is located in the home network of the MS to provide session control of multimedia services. The S-CSCF interacts with the application servers to obtain value added services. Furthermore, the S-CSCF communicates with the HSS and the AuC to receive IMS-related subscriber data and authentication information of the MS. An interrogating-CSCF (I-CSCF) is a firewall for the SIP messages toward the home network, and is responsible for selecting an S-CSCF for the MS.

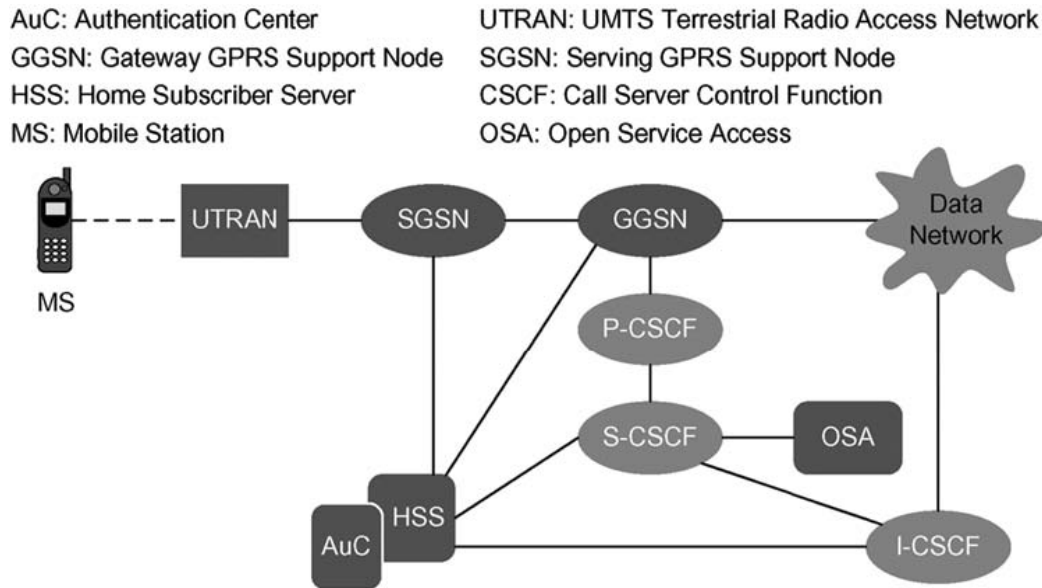
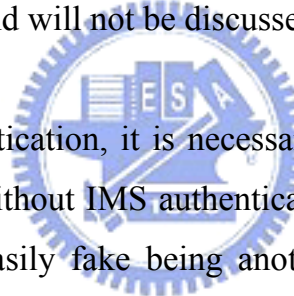


Fig. 3.1. UMTS architecture for packet switched service domain.

In UMTS, when an MS sends an “Initial L3 message” (e.g., location update request, connection management service request, routing area update request, attach request, paging response, etc.) to the SGSN, the SGSN may be triggered to authenticate the user. The authenticating parties are HSS/AuC in the home network and the universal subscriber identity module (USIM) in the MS. GPRS authentication consists of two major procedures [19], [23].

- **Distribution of authentication information from the AuC to the SGSN:** The SGSN sends an authentication data request to the HSS/AuC with the parameter international mobile subscriber identity (IMSI) of the MS, and receives a response with an array of authentication vectors (AVs) from the AuC. An authentication vector consists of a random number **RAND**, an expected response **XRES**, a cipher key **CK**, an integrity key **IK**, and an authentication token **AUTN**. Each AV is good for one authentication and key agreement between the SGSN and the MS.
- **Authentication and key agreement between the SGSN and the MS:** This

procedure performs authentication between an MS and the network by showing knowledge of a preshared secret key \mathbf{K} that is only available in the USIM of the MS and the AuC. The SGSN invokes the authentication procedure with an authentication vector. This procedure supports mutual authentication between the MS and the network. Specifically, the **AUTN** is used by the MS to authenticate the network, and the **RES/XRES** pair is used by the SGSN to authenticate the MS (where the **RES** is generated by the MS). Details of the procedure will be given in Section II-A. The MS also computes two keys **CK** and **IK** using the received **RAND** and the preshared secret key \mathbf{K} stored in the USIM. On the network side, the SGSN passes **CK** and **IK** to the UTRAN. During data transmission, **CK** and **IK** are used for ciphering and integrity between the MS and the UTRAN. Data ciphering and integrity is out of the scope of this chapter, and will not be discussed further.



In addition to GPRS authentication, it is necessary to authenticate the MS before it can access IMS services. Without IMS authentication, a mobile user who passes the GPRS authentication can easily fake being another IMS user. Details of the fake procedure will be elaborated in Section II-C. IMS authentication is performed between the IMS subscriber identity module (ISIM) in the MS and the AuC in the home network [24]. This procedure is basically the same as the GPRS authentication. In this procedure, the CSCF first sends a multimedia authentication request to the HSS/AuC with the IP multimedia private identity (IMPI) of the MS, and receives a response with an array of AVs. (This step is skipped if the CSCF already has the AV array.) The CSCF then invokes the IMS authentication and key agreement procedure with an authentication vector. The MS authenticates the network through the received **AUTN** and the CSCF authenticates the MS using the **RES/XRES** pair. Detailed message flow of this procedure will be given in Section II-B.

Although both GPRS and IMS authentications are necessary, most steps in these

two “authentication passes” are duplicated. In other words, the two-pass authentication proposed in 3GPP 33.203 [24] is not efficient. In this chapter, we propose a one-pass authentication procedure that effectively combines both the GPRS and the IMS authentications. We prove that this simplified one-pass authentication procedure correctly authenticate the IMS users.

3.2 3GPP TWO-PASS AUTHENTICATION

This section describes the 3GPP two-pass authentication procedure. We first describe GPRS authentication in Section 3.2-A, and then we elaborate more on IMS authentication in Section 3.2-B. In Section 3.2-C, we explain why authentication must be performed in both the GPRS and the IMS levels.

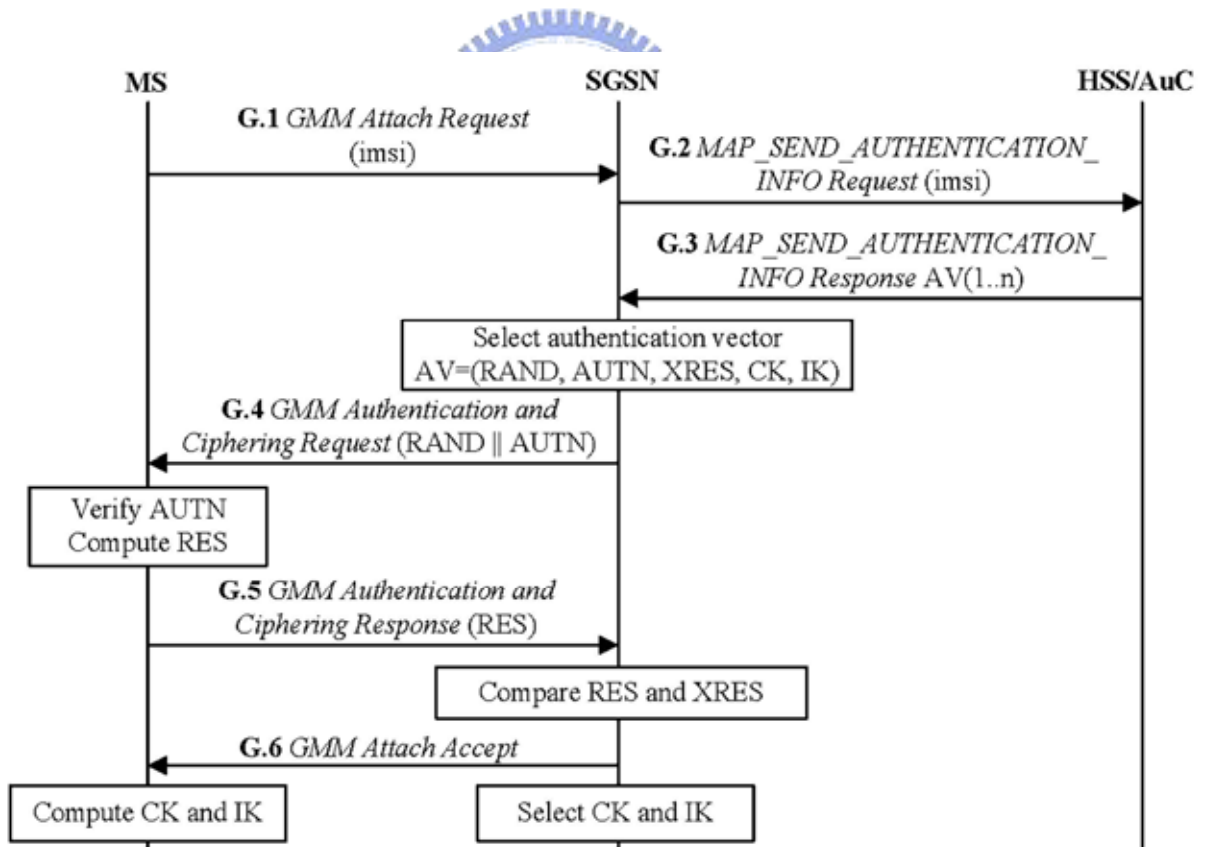


Fig. 3.2. Message flow for 3GPP GPRS authentication.

A. GPRS Authentication

When an MS invokes the GPRS access (e.g., turns on its power), the MS sends an attach request to the SGSN. This message will trigger the GPRS authentication [20], which is implemented by GPRS mobility management (GMM) between the MS and the SGSN, and signaling system number 7 (SS7) mobile application part (MAP) between the SGSN and the HSS/AuC [25]. This procedure consists of the following steps (see Fig. 3.2).

Step G.1) Consider an MS with the IMSI value *imsi* and the IMPI value *impi*. To access the GPRS services, the MS sends a **GMM Attach Request** (with the parameter **IMSI = *imsi***) to the SGSN.

Step G.2) If the SGSN has the AVs of the MS, then Steps G.2 and G.3 are skipped. Otherwise, the SGSN must obtain the AV's from the HSS/AuC. That is, the SGSN invokes the authentication vector distribution procedure by sending a **MAP SEND AUTHENTICATION INFO Request** message to the HSS/AuC (with the parameter **IMSI = *imsi***).

Step G.3) The HSS/AuC uses *imsi* to retrieve the record of the MS, and generates an ordered array of AVs (based on the preshared secret key **K** in the MS record). The generated AV array is sent to the SGSN through a **MAP SEND AUTHENTICATION INFO Response** message.

Step G.4) The SGSN selects the next unused authentication vector in the ordered AV array and sends the parameters **RAND** and **AUTN** (from the selected authentication vector) to the MS through a **GMM Authentication and Ciphering Request** message.

Step G.5) The MS checks whether the received **AUTN** can be accepted. If so, it produces a response **RES** that is sent back to the SGSN through a **GMM**

Authentication and Ciphering Response message. The SGSN compares the received **RES** with the **XRES**. If they match, then the authentication and key agreement exchange is successfully completed.

Step G.6) The SGSN sends a **GMM Attach Accept** message to the MS, and the attach procedure is completed.

After GPRS authentication, GPRS registration follows (details of GPRS registration can be found in [27]). Then, the MS performs packet data protocol (PDP) context activation to obtain access to the GPRS network. The PDP context specifies the application-layer packet data protocol and the routing information used for the GPRS communication session (see [28] for the details).

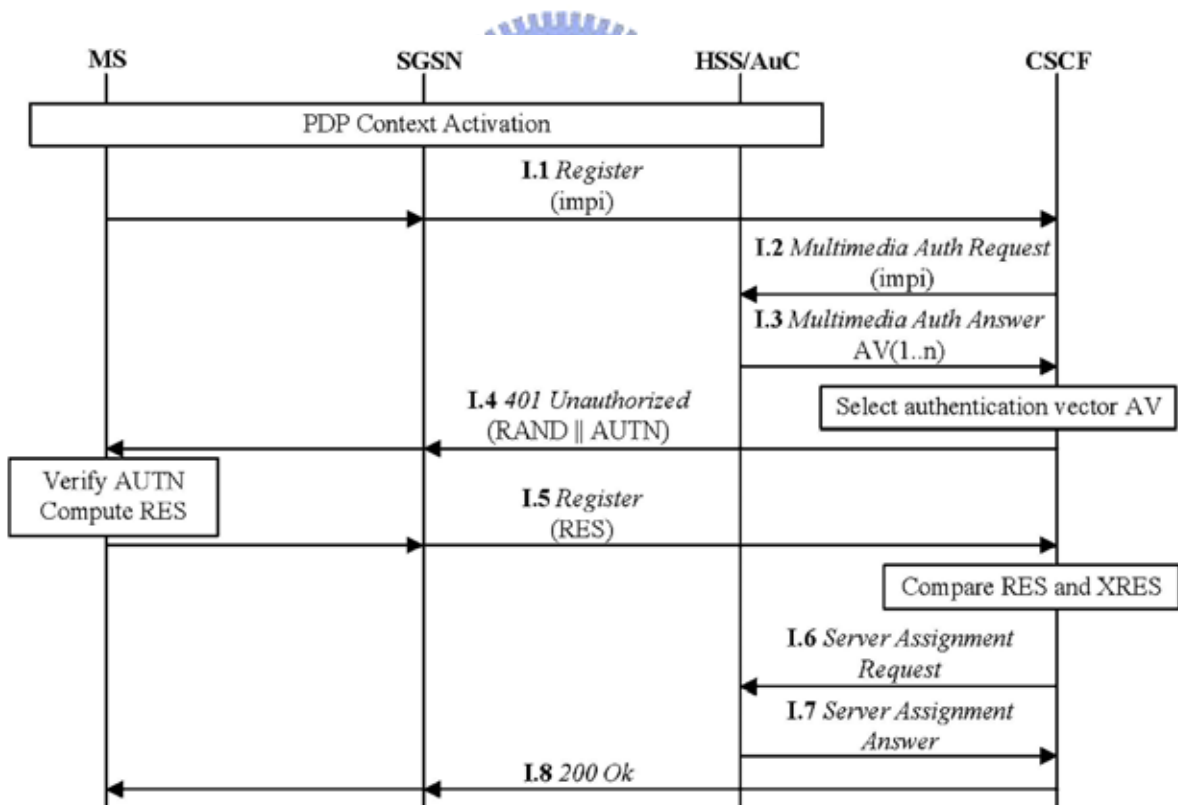


Fig. 3.3. Message flow for 3GPP IMS authentication.

B. IMS Authentication

After PDP context activation, the MS can request the IMS services through the registration procedure illustrated in Fig. 3.3. In this procedure, the MS interacts with the S-CSCF possibly through P-CSCF and I-CSCF. To simplify our discussion, Fig. 3 uses the term “CSCF” to represent the proxy, interrogating, and service functions of CSCF. Details of message exchanges among these CSCFs are given in [28]. IMS authentication/reg-istration is implemented by SIP and Cx protocols [29], [30], which consists of the following steps.

Step I.1) The MS sends a SIP Register message to the CSCF (with the parameter $IMPI = impi$) through the SGSN.

Step I.2) Assume that the CSCF does not have the AVs for the MS. The CSCF invokes the authentication vector distribution procedure by sending a Cx Multimedia Authentication Request message to the HSS/AuC (with the parameter $IMPI = impi$).

Step I.3) The HSS/AuC uses $impi$ to retrieve the record of the MS, and generate an ordered array of AVs. The HSS/AuC sends the AV array to the CSCF through a Cx Multimedia Authentication Answer message.

Step I.4) The CSCF selects the next unused authentication vector from the ordered AV array and sends the parameters **RAND** and **AUTN** (from the selected authentication vector) to the MS through a SIP 401 Unauthorized message.

Step I.5) The MS checks whether the received **AUTN** can be accepted. If so, it produces a response **RES**. The MS sends this response back to the CSCF through a SIP Register message. The CSCF compares the received **RES** with the **XRES**. If they match, then the authentication and key agreement exchange is successfully completed.

Step I.6) The CSCF sends a Cx Server Assignment Request message to the

HSS/AuC.

Step I.7) Upon receipt of the Server Assignment Request, the HSS/AuC stores the CSCF name and replies a Cx Server Assignment Answer message to the CSCF.

Step I.8) The CSCF sends a 200 ok message to the MS through the SGSN, and the IMS registration procedure is completed.

In the above procedure, Steps I.1–I.5 exercise authentication, and Steps I.6–I.8 perform registration.

TABLE 3.1 IDENTICAL STEPS IN GPRS AND IMS AUTHENTICATIONS

GPRS authentication (SS7 MAP)	IMS authentication (SIP/Cx)
G.2: MAP_SEND_AUTHENTICATION_INFO Request Parameter: IMSI	I.2: Multimedia Authentication Request Parameter: IMPI
G.3: MAP_SEND_AUTHENTICATION_INFO Response Parameter: AV[1..n]	I.3: Multimedia Authentication Answer Parameter: AV[1..n]
G.4: User Authentication Request Parameter: RAND AUTN	I.4: 401 Unauthorized Parameter: RAND AUTN
G.5: User Authentication Response Parameter: RES	I.5: Register Parameter: RES
G.6: GMM Attach Accept	I.8: 200 Ok

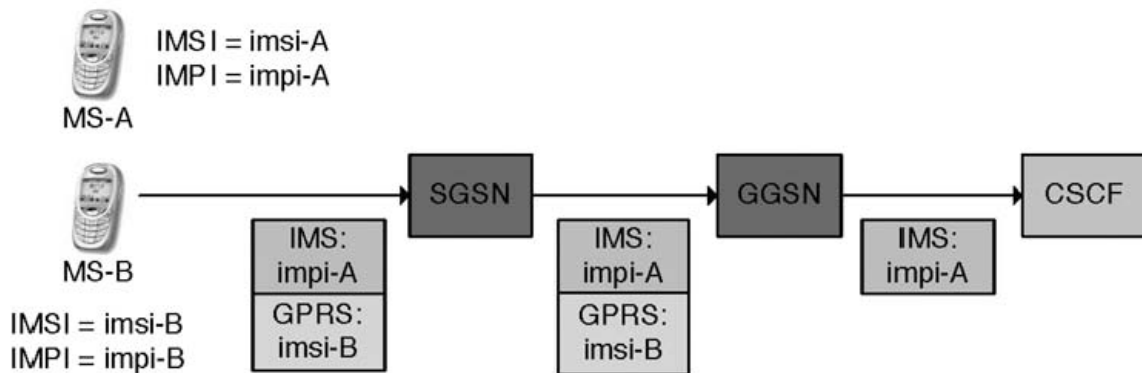


Fig. 3.4. Illegal IMS registration.

C. Fraudulent IMS Usage

Although GPRS authentication is implemented by GMM and SS7 MAP, and IMS authentication is implemented by SIP and Cx, many steps of these two authentication procedures are duplicated (see Table 3.1). Unfortunately, these redundant steps are required. That is, after GPRS authentication, it is necessary to authenticate the MSs again at the IMS level. Without IMS authentication, an IMS user may pretend to be another IMS user. Consider the example in Fig. 3.4, where there are two MSs. MS-A has the IMSI value *imsi-A* and the IMPI value *impi-A*. MS-B has the IMSI value *imsi-B* and the IMPI value *impi-B*. Suppose that MS-B is a legal GPRS user and has passed the GPRS authentication (by using *imsi-B*) to obtain GPRS network access. If no IMS authentication is required, MS-B may perform IMS registration by sending the CSCF a Register request that includes the MS-A's IMPI value *impi-A* as a parameter. The CSCF will consider this IMS registration as a legal action activated by MS-A. Therefore, MS-B can illegally access the IMS services of MS-A. The above example shows that IMS-level authentication is required to prevent illegal access to the IMS services. In the next section, we describe an one-pass authentication procedure for both GPRS and IMS authentications. Our approach

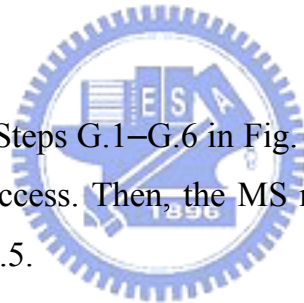
significantly reduces the number of accesses to the HSS/AuC.

3.3 ONE-PASS AUTHENTICATION PROCEDURE

This section proposes an one-pass authentication (performed at the GPRS level) that can authenticate an IMS user without explicitly performing the IMS-level authentication. In our approach, the SGSN implements a SIP application level gateway (ALG) [31] that modifies the format of SIP messages (to be elaborated). We first describe the SIP message flow of the one-pass procedure. Then, we provide a brief cost comparison between the one-pass and the two-pass procedures.

A. SIP Message Flow

After GPRS authentication (Steps G.1–G.6 in Fig. 3.2) the MS performs PDP context activation to obtain GPRS access. Then, the MS registers to the IMS through Steps I*.1–I*.4 illustrated in Fig. 3.5.



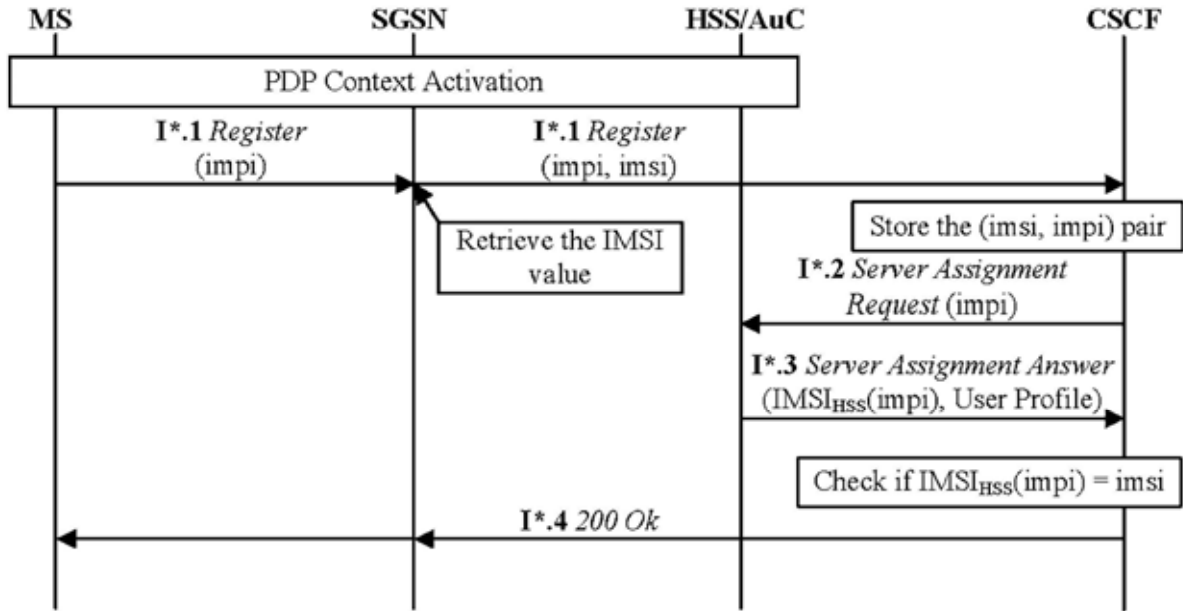


Fig. 3.5. IMS registration (one-pass authentication).

Step I*.1) The MS sends a SIP Register message to the SGSN with the parameter $IMPI = impi$. Note that after PDP context activation, the SGSN can identify the IMSI of the MS that transmits the GPRS packets [20]. The SIP ALG in the SGSN adds the IMSI value (i.e., $imsi$) of the MS in the Register message and forwards it to the CSCF. Details of a possible SIP ALG implementation can be found in [31].

Step I*.2) The CSCF stores the $(imsi, impi)$ pair in the MS record, and sends a Cx Server Assignment Request message to the HSS/AuC with the parameter $IMPI = impi$. We note that if the CSCF has stored the $(imsi, impi)$ pair before, then Steps I*.2 and I*.3 are skipped.

Step I*.3) The HSS/AuC uses the received IMPI value $impi$ as an index to retrieve the IMSI and the user profile of the MS. We denote $IMSI_{HSS}(impi)$ as the IMSI value retrieved from the HSS/AuC. The HSS/AuC stores the CSCF name and sends a Cx Server Assignment Answer to the CSCF (with the parameters $IMSI_{HSS}(impi)$ and user profile).

Step I*.4) The CSCF checks whether the value $imsi$ and $IMSI_{HSS}(impi)$ are the same. If so, the CSCF sends a SIP 200 Ok message to the SGSN and the authentication is considered successful. If $IMSI_{HSS}(impi) \neq imsi$, then it implies that the registration is illegal (i.e., the scenario illustrated in Fig. 4 occurs). Suppose that $IMSI_{HSS}(impi) = imsi$. The SGSN forward the 200 Ok message to the MS, and the IMS registration procedure is successfully completed.

TABLE 3.2 COMPARING THE ONE-PASS AND THE TWO-PASS AUTHENTICATION PROCEDURES IN IMS REGISTRATION

One-Pass Procedure	Two-Pass Procedure
I*.1: Register Parameters: $impi$ and $imsi$	I.1: Register Parameter: $impi$
-	I.2: Multimedia Authentication Request Parameter: $impi$
-	I.3: Multimedia Authentication Answer Parameter: AV[1..n]
-	I.4: 401 Unauthorized Parameter: RAND AUTN
-	I.5: Register Parameter: RES
I*.2: Server Assignment Request	I.6: Server Assignment Request
I*.3: Server Assignment Response	I.7: Server Assignment Response
I*.4: 200 Ok	I.8: 200 Ok

B. Cost Analysis

Table II compares the steps executed in the one-pass and the two-pass authentication procedures. Suppose that the expected SIP message delivery cost between an MS and the CSCF is one unit, and the expected Cx message delivery cost

between the CSCF and the HSS/AuC is α units. It is anticipated that for the following two reasons.

- The CSCF and the HSS/AuC exchange the Cx messages through IP network. On the other hand, besides the IP network overhead, SIP communications between the MS and the CSCF involves GPRS core network and UTRAN radio network.
- The CSCF and the AuC/HSS are typically located at the same location, while the MS is likely to reside at a remote location.

It is clear that the expected IMS registration C_1 for the one-pass procedure (see Fig. 3.5) is

$$C_1 = 2 + 2\alpha. \quad (3.1)$$

Note that Step I*.1 needs to trigger SIP ALG for SIP message analysis. Since this action is executed in micro kernel of the SGSN, the overhead can be ignored as compared with SIP message exchange. Similarly, the extra cost of $\text{IMSI}_{\text{HSS}}(\text{impi})$ and imsi comparison at Step I*.4 can be ignored. Our analysis assumes that the $(\text{imsi}, \text{impi})$ pair does not exist at Step I*.1. Therefore Steps I*.2 and I*.3 are always executed. This assumption favors the two-pass procedure.

In the two-pass procedure, if the distribution of authentication vectors from the HSS/AuC to the SGSN (Steps I.1–I.4 in Fig. 3) is performed, then the expected IMS registration cost $C_{2,1}$ is expressed as

$$C_{2,1} = 4 + 4\alpha. \quad (3.2)$$

If the authentication vector distribution is not executed in the two-pass procedure, then the expected IMS registration cost $C_{2,2}$ is expressed as

$$C_{2,2} = 4 + 2\alpha. \quad (3.3)$$

Like periodic location update in UMTS [32], IMS registration is periodically

performed. In **Steps I.2** and **I.3** of the two-pass procedure, an AV array of size n (where $n \geq 1$) is sent from the HSS/AuC to the CSCF. Therefore, one out of the n IMS registrations incurs execution of **Steps I.2** and **I.3**. Therefore, from (3.2) and (3.3), the expected IMS registration cost C_2 for the two-pass procedure is

$$C_2 = \left(\frac{1}{n}\right) C_{2,1} + \left(\frac{n-1}{n}\right) C_{2,2} = 4 + \left(\frac{n+1}{n}\right) 2\alpha \quad (3.4)$$

From (3.1) and (3.4), the improvement S of the one-pass procedure over the two-pass procedure is

$$S = \frac{C_2 - C_1}{C_2} = \frac{n + \alpha}{2n + (n+1)\alpha} \quad (3.5)$$

Figure 3.1 plots S as a function of n and α . The figure indicates that the one-pass procedure can save up to 50% of the SIP/Cx traffic for IMS registration/authentication, as compared with the two-pass procedure. Another significant advantage of the one-pass procedure is that it consumes much less AVs (about 50% less) than the two-pass procedure.

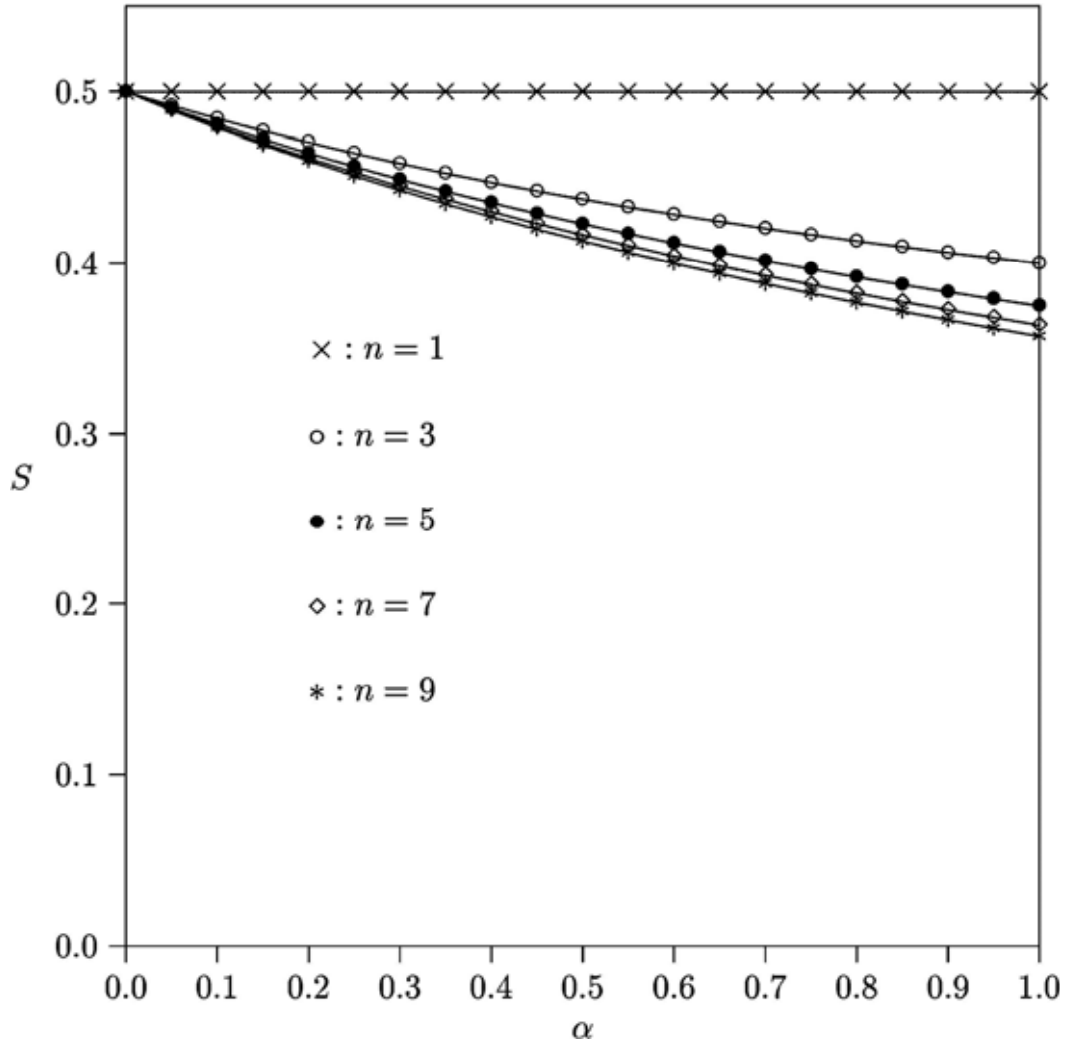


Figure 3.1. Improvement of the One-Pass Procedure over the Two-Pass Procedure

One may argue that implementation of a SIP ALG is required in the one-pass procedure. Since IMS is based on SIP, a SIP ALG is required for other purposes (see an example in [26]). Therefore, the one-pass procedure will incur little extra cost for implementing SIP ALG.

3.4 Correctness of The One-Pass Procedure

In this section, we prove that the one-pass authentication procedure correctly authenticates the IMS users. In UMTS, every MS maintains the attributes IMSI, IMPI,

and the preshared secret key \mathbf{K} in its SIM card. Consider an MS with $\text{IMSI} = \text{imsi}$, $\text{IMPI} = \text{impi}$, and $\mathbf{K} = k$. To simplify our discussion, we assume that these parameters are grouped into a set $R_{\text{MS}} = \{\text{imsi}, \text{impi}, k\}$ in the SIM card of the MS. Define functions IMSI_{MS} , IMPI_{MS} , and \mathbf{K}_{MS} such that for any $x \in R_{\text{MS}}$

$$\text{IMSI}_{\text{MS}}(x) = \text{imsi}, \text{ where } \text{imsi} \text{ is the IMSI value in } R_{\text{MS}}. \quad (3.6)$$

$$\text{IMPI}_{\text{MS}}(x) = \text{impi}, \text{ where } \text{impi} \text{ is the IMPI value in } R_{\text{MS}}. \quad (3.7)$$

$$\mathbf{K}_{\text{MS}}(x) = k, \text{ where } k \text{ is the } \mathbf{K} \text{ value in } R_{\text{MS}}. \quad (3.8)$$

Based on the above definitions, it is clear that, for example

$$\text{IMSI}_{\text{MS}}(\text{impi}) = \text{IMSI}_{\text{MS}}(k) = \text{imsi}.$$

Similarly, for every MS, the HSS/AuC maintains a record R_{HSS} that consists of attributes IMSI, IMPI, and \mathbf{K} . That is, for an MS who has legal GPRS and IMS accesses

$$R_{\text{HSS}} = \{\text{imsi}, \text{impi}, k\} = R_{\text{MS}}.$$

Like (3.6)-(3.8), we define functions IMSI_{HSS} , IMPI_{HSS} , and \mathbf{K}_{HSS} such that for any $x \in R_{\text{HSS}}$,

$$\text{IMSI}_{\text{HSS}}(x) = \text{imsi}, \text{ where } \text{imsi} \text{ is the IMSI value in } R_{\text{HSS}}. \quad (3.9)$$

$$\text{IMPI}_{\text{HSS}}(x) = \text{impi}, \text{ where } \text{impi} \text{ is the IMPI value in } R_{\text{HSS}}. \quad (3.10)$$

$$\mathbf{K}_{\text{HSS}}(x) = k, \text{ where } k \text{ is the } \mathbf{K} \text{ value in } R_{\text{HSS}}. \quad (3.11)$$

In 3G 23.060 [20] and 3G 33.203 [24], MS authentication at the GPRS and the IMS levels are based on the following Theorem.

Theorem 1: Suppose that an MS claims that it has the IMSI value imsi and the IMPI value impi . Then,

- a) The MS is a legal GPRS user if $\mathbf{K}_{\text{MS}}(\text{imsi}) = \mathbf{K}_{\text{HSS}}(\text{imsi})$.
- b) The MS is a legal IMS user if $\mathbf{K}_{\text{MS}}(\text{impi}) = \mathbf{K}_{\text{HSS}}(\text{impi})$.

Note that Theorem 1 does not hold if an illegal user already possesses the SIM information of a legal user (e.g., by duplicating the SIM card through the SIM card reader [26]). This issue was addressed in [33]. In this chapter, we assume that such fraudulent usage does not occur. 3GPP GPRS authentication procedure (i.e., Steps G.1–G.6) checks if both a GPRS user and the HSS/AuC have the same preshared secret key \mathbf{K} using Theorem 1 and Fact 1a below. Similarly, 3GPP IMS authentication procedure (i.e., Steps I.1–I.8) checks if both an IMS user and the HSS/AuC have the same preshared secret key using Theorem 1 and Fact 1b.

Fact 1:

- a) For an MS claiming $\text{IMSI} = imsi$, if $\text{XRES} = \text{RES}$, then $\text{K}_{\text{MS}}(imsi) = \text{K}_{\text{HSS}}(imsi)$.
- b) For an MS claiming $\text{IMPI} = impi$, if $\text{XRES} = \text{RES}$, then $\text{K}_{\text{MS}}(impi) = \text{K}_{\text{HSS}}(impi)$.

Now, we prove that the one-pass authentication correctly authenticates the IMS users (i.e., the one-pass procedure checks if $\text{K}_{\text{MS}}(impi) = \text{K}_{\text{HSS}}(impi)$). From the definitions of the IMSI_{HSS} and K_{HSS} functions [i.e., (3.9) and 3.11)], it is trivial to have the following fact.

Fact 2:

For any IMPI value $impi$, if $\text{IMSI}_{\text{HSS}}(impi) = imsi$, then $\text{K}_{\text{HSS}}(impi) = \text{K}_{\text{HSS}}(imsi)$.

With Fact 2, correctness of the one-pass authentication procedure is guaranteed according to the following two theorems.

Theorem 2: Suppose that

- a) an MS with the IMSI value $imsi$ has passed the GPRS authentication; that is

$$\text{K}_{\text{MS}}(imsi) = \text{K}_{\text{HSS}}(imsi). \quad (3.12)$$
- b) The MS claims that its IMPI value is $impi$.
- c) The network maps $impi$ to the IMSI value $imsi$; that is

$$\text{IMSI}_{\text{HSS}}(\text{impi}) = \text{imsi}. \quad (3.13)$$

Then, the MS is a legal IMS user. In other words

$$\text{K}_{\text{MS}}(\text{impi}) = \text{K}_{\text{HSS}}(\text{impi}). \quad (3.14)$$

Proof:

From hypothesis a, $\text{imsi} \in \text{R}_{\text{MS}}$. In hypothesis b, the MS claims that it has the IMPI value impi , which implies that $\text{impi} \in \text{R}_{\text{MS}}$. From (3.8)

$$\text{K}_{\text{MS}}(\text{imsi}) = \text{K}_{\text{MS}}(\text{impi}). \quad (3.15)$$

From Fact 2 and (3.13) in hypothesis c, we have

$$\text{K}_{\text{HSS}}(\text{impi}) = \text{K}_{\text{HSS}}(\text{imsi}). \quad (3.16)$$

From (3.12) in hypothesis a and (3.16), we have

$$\text{K}_{\text{MS}}(\text{imsi}) = \text{K}_{\text{HSS}}(\text{impi}). \quad (3.17)$$

From (3.15) and (3.17), we have

$$\text{K}_{\text{MS}}(\text{impi}) = \text{K}_{\text{HSS}}(\text{impi}).$$

In other words, if hypotheses a–c hold, an MS is a legal IMS user with $\text{IMPI} = \text{impi}$.

Q.E.D.

Theorem 3: The one-pass authentication procedure correctly authenticates the IMS users; that is, for an MS claiming the IMPI value impi , the one-pass procedure recognizes the MS as a legal IMS user if $\text{K}_{\text{MS}}(\text{impi}) = \text{K}_{\text{HSS}}(\text{impi})$.

Proof:

After Steps G.1–G.6 have been executed, the network verifies that $\text{K}_{\text{MS}}(\text{imsi}) = \text{K}_{\text{HSS}}(\text{imsi})$; i.e., (3.12) in Theorem 2 is satisfied.

At Step I*.1, the MS claims that its IMPI value is impi and, therefore, the network assumes that $\text{K}_{\text{MS}}(\text{imsi}) = \text{K}_{\text{MS}}(\text{impi})$; i.e., (3.15) in Theorem 2 is satisfied.

At Step I*.4, the one-pass authentication checks if $\text{IMSI}_{\text{HSS}}(\text{impi}) = \text{imsi}$ [i.e., (3.13) in Theorem 2 is checked]. If so, $\text{K}_{\text{MS}}(\text{impi}) = \text{K}_{\text{HSS}}(\text{impi})$ as a direct consequence of Theorem 2, and the authentication procedure recognizes the MS as a legal user

(according to Theorem 1). Otherwise, the authentication fails.

In other words, the one-pass procedure follows Theorem 1 to authenticate an MS.

Q.E.D.

3.5 Summary

This chapter proposed an efficient IMS registration procedure without explicitly performing tedious authentication steps. As specified by the 3GPP, after a UMTS mobile user has obtained GPRS network access through GPRS authentication, the “same” authentication procedure must be executed again at the IMS level (during IMS registration) before it can receive the IP multimedia services. This chapter described an one-pass authentication procedure, which only needs to perform GPRS authentication. At the IMS registration, the one-pass procedure performs several simple operations to verify if a user is legal. We prove that the one-pass procedure correctly authenticates the IMS users. Compared with the eight-step two-pass authentication, the four-step one-pass authentication saves two to four SIP/Cx message exchanges among the MS, the SGSN, the CSCF, and the HSS/AuC. Our study indicates that this new approach can save up to 50% of the network traffic generated by the IMS registration. This approach also saves 50% of the storage for buffering the authentication vectors.

Chapter 4

A SIP-based Call Center with Waiting Time Prediction

A call center in a company is a centralized office that answers incoming telephone calls from customers or makes outgoing telephone calls to customers. A call center may also provide fax, Voice over IP (VoIP), e-mail or web-based interactions. This chapter proposes a plug-in modular architecture for an Internet call center with waiting time prediction. We implement this plug-in module in the SIP Express Router (SER) platform. Then we propose two output measures and develop a discrete event simulation model to investigate the performance of the waiting time prediction algorithm for the call center. Our study indicates that the waiting times can be more accurately predicted when the call arrival rate is large.

4.1 INTRODUCTION

A call center [34] in a company is a centralized office that answers incoming telephone calls from customers or makes outgoing telephone calls to customers. A call center may also provide fax, Voice over IP (VoIP), e-mail or web-based interactions. In a call center, when all agents are busy, the arriving customer calls have to wait in a queue. A waiting customer will be served after an agent becomes available, or abandon the call request before any agent is available. To improve

customer satisfaction, it is essential to inform the waiting customers of anticipated delays before they can be served.

Session Initiation Protocol (SIP) [6, 35] is a text-based protocol for initiating communication sessions between users. These sessions may include voice, video calls, multimedia conferencing, streaming media services, games, and so on. SIP utilizes Real-time Transport Protocol (RTP) [2, 36] to transmit real-time packets over the Internet.

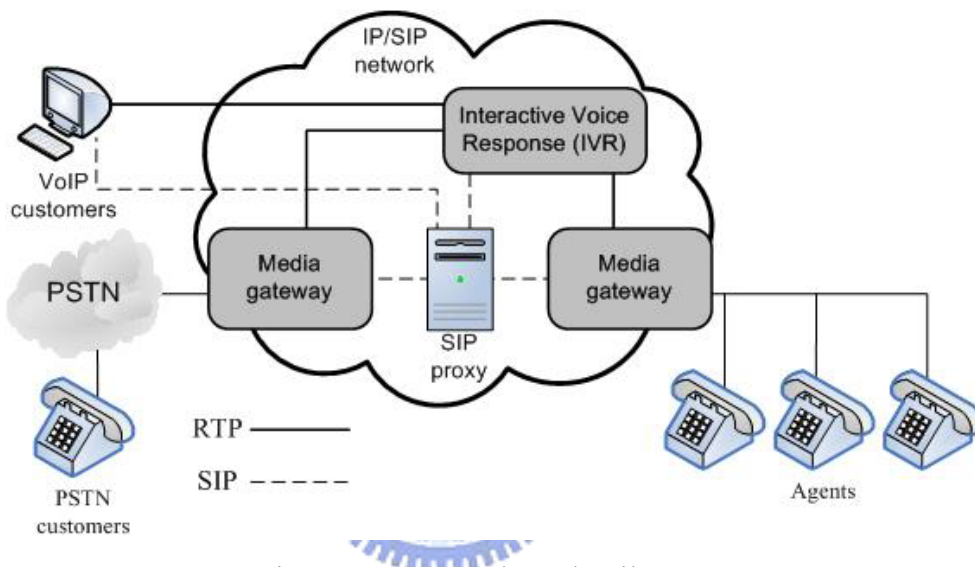


Figure 4.1. VoIP-based call center.

Figure 4.1 shows a VoIP-based call center [37, 38], where the customers access the call center through telephones in the PSTN or IP-based devices in the Internet. The IP/SIP network interfaces to the PSTN through border elements such as media gateways. The media gateways recognize Dual Tone Multi-Frequency (DTMF) [39] signals from customers and convert them to SIP messages or RTP media packets. The VoIP-based call center which contains a SIP proxy provides call control functions such as call transfer, conferencing, and Automatic Call Distribution (ACD).

This chapter proposes a plug-in modular architecture for VoIP-based call center with waiting time prediction. This plug-in module is implemented in the SIP Express Router (SER) platform [40]. In Section 2, the SER platform and the plug-in module

are elaborated. In Section 3, the message flows of the call center are presented. The proposed waiting time prediction algorithm and performance analysis are described in Sections 4. A callback mechanism is described in Section 5.

4.2 Automatic Call Distributor Module for SER

The SIP Express Router (SER) [40] is a high-performance, configurable, free SIP server. It can act as a SIP registrar, proxy or redirect server. SER provides an application-server interface, presence support, SMS gateway, RADIUS/syslog accounting and authorization, etc. Written in the C language, the SER can be ported on Linux, BSD and Solaris.

The SER defines a flexible interface for add-on function modules. These add-on modules are compiled and stored as shared objects. When the SER is instructed to load a module using “loadmodule” command in the configure file, it opens and loads the specified module file. The exported symbols and functions are then registered in this module. These functions are executed by the SER in a script called “route block” that defines the rules to route and modify SIP messages.

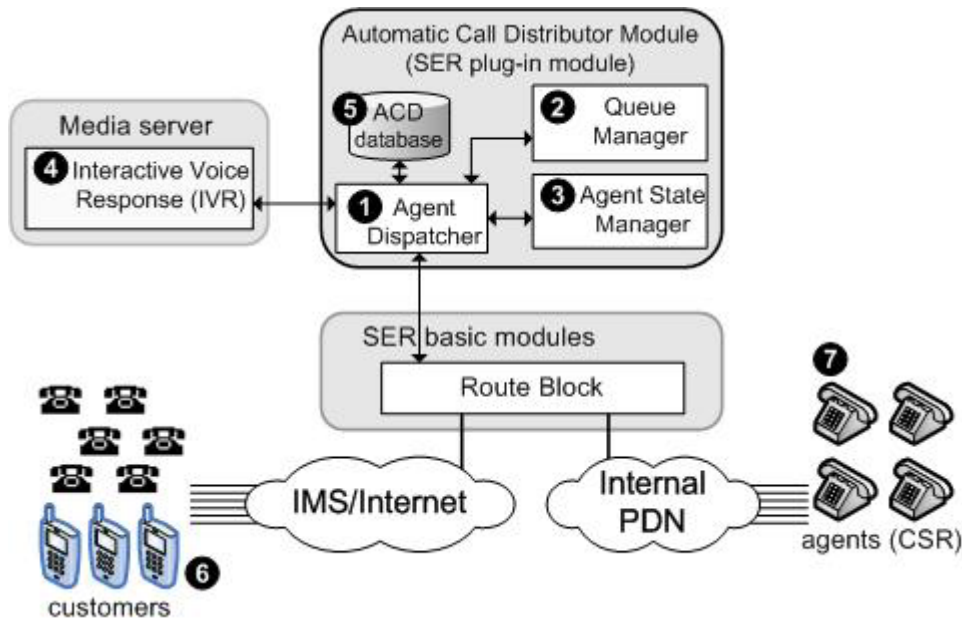


Figure 4.2. The proposed call center architecture.

Figure 4.2 depicts the plug-in SER module called *Automatic Call Distributor* (ACD) module for the proposed call center system. In this system, a customer (Figure 4.2 (6)) connects to the call center by dialing the call center's identity, which is a SIP Uniform Resource Identifier (URI; e.g., 0941@nctu.edu.tw). The call center interacts with the ACD to select an available agent. In the call center, the agents handling inbound calls are referred to as the *Customer Service Representatives* (CSRs; see Figure 4.2 (7)). The ACD database (Figure 4.2 (5)) maintains a record to store call information related to each of the agents, such as agent SIP URI, group name, agent state, Call-ID of the current call handled by this agent, customer's SIP URI, average service time, service time of the last call, and so on.

The *agent dispatcher* (Figure 4.2 (1)) dispatches agents to handle the incoming calls. If all agents are busy, the agent dispatcher interacts with the *queue manager* (Figure 4.2 (2)) to predict the waiting time and notifies the Interactive Voice Response (IVR; see Figure 4.2 (4)) to announce the predicted waiting time to the customer. The queue manager supports queue management and calculation of predicted waiting times. It keeps a table which records the waiting customer's SIP

URI, call arrival time, predicted waiting time, Call-ID, and so on. The prediction algorithm will be elaborated in Section 4. The *agent state manager* (Figure 4.2 (3)) maintains a finite state machine for each agent and interacts with the route block through agent dispatcher using function calls. This manager treats the SIP messages as events and decides how to react depending on the agent's current state, such as TALKING, PENDING, IDLE, ERROR, etc (see Figure 4.3).

The Call Center provides the following interfaces:

- The interface between the SER and ACD module: The ACD module provides four functions to be invoked by the SER route block. The `is_user_in()` function checks if a SIP URI is a member of the specified group. The `select_agent()` function (Figure 4.4) selects an available call agent to deal with an inbound call. This function also rewrites the SIP URI in the INVITE message to a new SIP URI of the selected call agent. The `rewrite_request_uri()` function is responsible for rewriting the SIP URI in the consecutive SIP messages. The `set_agent_state()` function (Figure 4.5) changes the call agent's state based on the SIP message type of the arrival message and the call agent's current state.
- The interface between the SER and the IVR: If the call center cannot find an available agent to handle an inbound call, the SER will route this call to the IVR module of the media server for voice announcement.
- The interface between the agent dispatcher and the IVR: The agent dispatcher provides call related information to the IVR such as predicted call waiting time, call identity, customer's SIP URI, and so on.
- The interface between the agent dispatcher and the ACD database: The ACD database provides customer related information such as customer service level, customer's preferred language and most familiar agent, the history of the customer's previous calls, and so on. The agent dispatcher utilizes this information to select a favorable agent for an inbound call.

- The interface between the IVR and the ACD database: The IVR stores the result of customer's input during interaction with a customer. The IVR queries the ACD database to obtain the customer id and related information to verify the correctness of the customer's input (e.g., credit card number).
- The interface between the agent state manager and the agent dispatcher: The agent state manager stores an agent's current state (see Figure 4.3 for state transition diagram). The agent dispatcher queries the agent state manager to select an available agent (with the IDLE state), and set the agent's new state through the function `set_agent_state()`.

The interface between the agent dispatcher and the queue manager: If the agent dispatcher cannot identify an available agent for an inbound call, the queue manager will buffer this customer request in the queue, calculates the predicted waiting time for this call. The waiting request is removed from the queue for service when an agent becomes available.



4.3 Call Center Message Flows

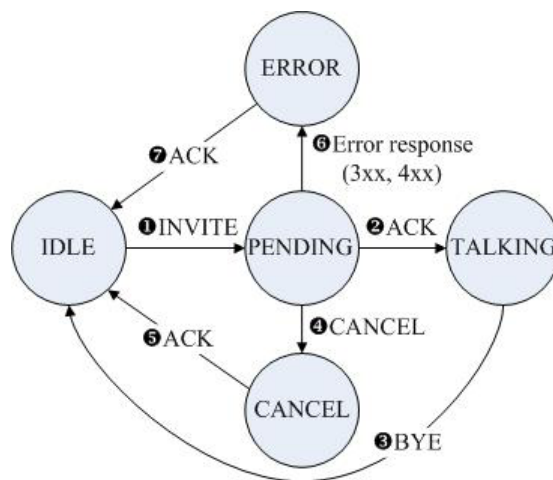


Figure 4.3. The FSM State Transition Diagram of ACD.

The ACD maintains a status record for each of the agents. The record is accessed by the agent state manager. A five-state Finite State Machine (FSM) is associated with the record with the following states (see Figure 4.3):

State IDLE: The agent is not serving any customer.

State PENDING: The agent has received the SIP INVITE message. The agent sends back a SIP 200 OK message if the agent accepts the call.

State TALKING: The agent has received the SIP ACK message. The media connection is established.

State CANCEL: The agent has received or originated the SIP CANCEL message. The recipient of the message sends back a SIP 200 OK response.

State ERROR: The agent has encountered an error and sent back a final response.

The transitions of the FSM are given below:

Transition 1: An incoming SIP INVITE message arrives at State IDLE. The agent sends back a SIP 180 Ringing response, and the state is changed to PENDING.

Transition 2: An incoming SIP ACK message arrives at State PENDING. The agent establishes the media connection with the customer, and the state is changed to TALKING.

Transition 3: A SIP BYE message was originated by either the agent or the customer at State TALKING. The agent terminates the media connection with the customer, and the state is changed to IDLE.

Transition 4: A SIP CANCEL message was originated by either the agent or the customer at State PENDING. The agent stops the ring tone, and the state is changed to CANCEL.

Transition 5: A SIP ACK message was originated by either the agent or the customer at State CANCEL. The state is changed to IDLE.

Transition 6: A final response with error code (e.g., 415 Unsupported Media Type) was sent by the agent at State PENDING. The agent stops the ring tone, and the state is changed to ERROR.

Transition 7: An incoming SIP ACK message arrives at State ERROR. The state is changed to IDLE.

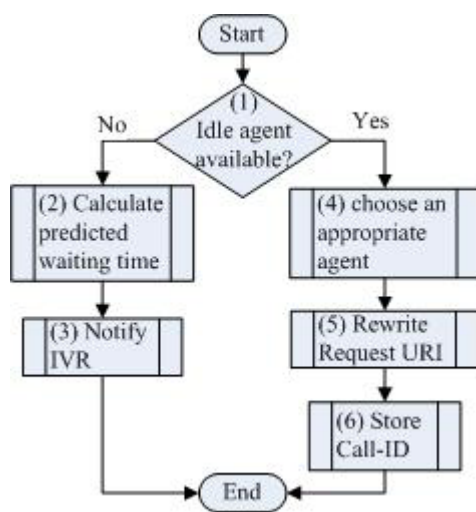


Figure 4.4. The select_agent() function.

Figure 4.4 shows the select_agent() function. At Step 1, if there are idle agents, the function choose one idle agent with the longest idle period at Step 4. The Request URI is rewritten to the agent's URI at Step 5 and the Call-ID is stored at Step 6. Otherwise (all agents are busy at Step 1), the predicted waiting time is calculated at Step 2 and the IVR is notified at Step 3.

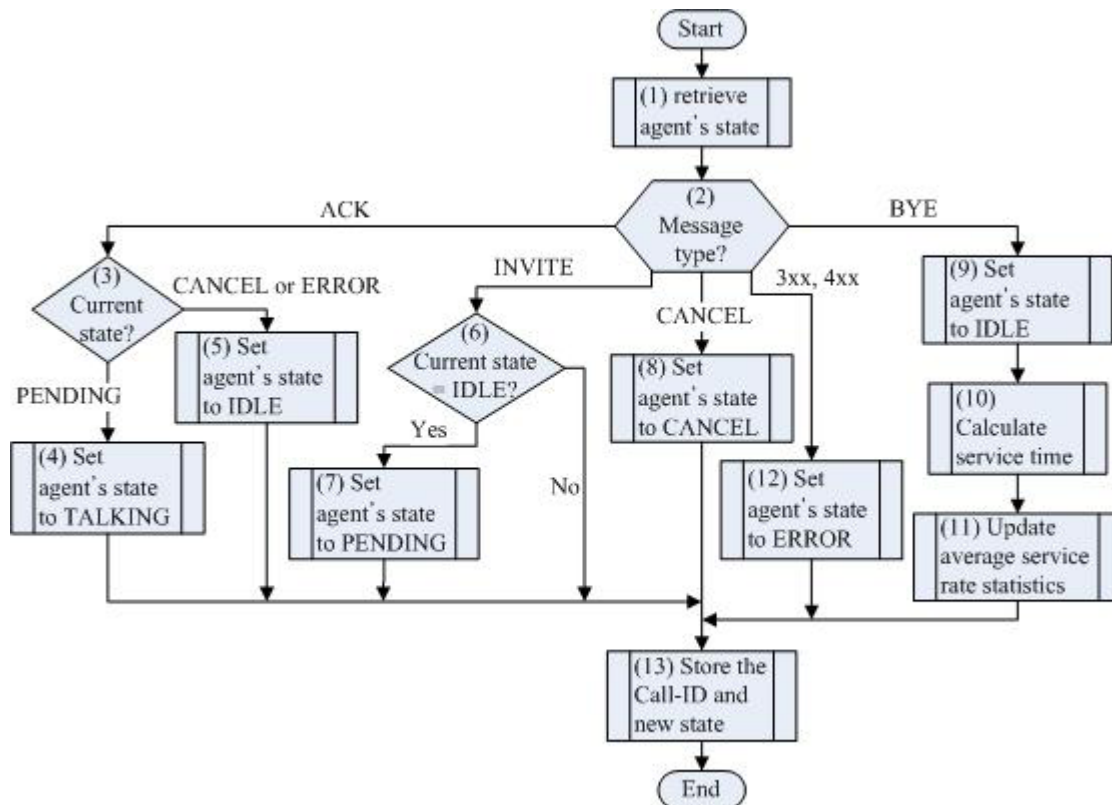


Figure 4.5. The set_agent_state() function.

Figure 4.5 shows the set_agent_state() function. Step 1 retrieves the agent's state from the agent state manager. At Step 2, the message is processed based on the message type described as follows.

Message=ACK: At Step 3, if the agent's current state is PENDING, then the state is changed to TALKING at Step 4. If the agent's current state is CANCEL or ERROR, then the state is changed to IDLE at Step 5.

Message=INVITE: At Step 6, if the agent's current state is IDLE, then the state is changed to PENDING at Step 7. Otherwise, no action is taken.

Message=CANCEL: At Step 8, the agent's state is set to CANCEL.

Message=3xx or 4xx: At Step 12, the agent's state is set to ERROR.

Message=BYE: At Step 9, the agent's state is set to IDLE. The service time for this call is calculated at Step 10. The average serve rate is updated at Step 11.

After the message has been processed, Step 13 is executed to store the new state of the agent at the ACD database.

The normal call setup procedure is illustrated in Figure 4.6, which consists of the following steps:

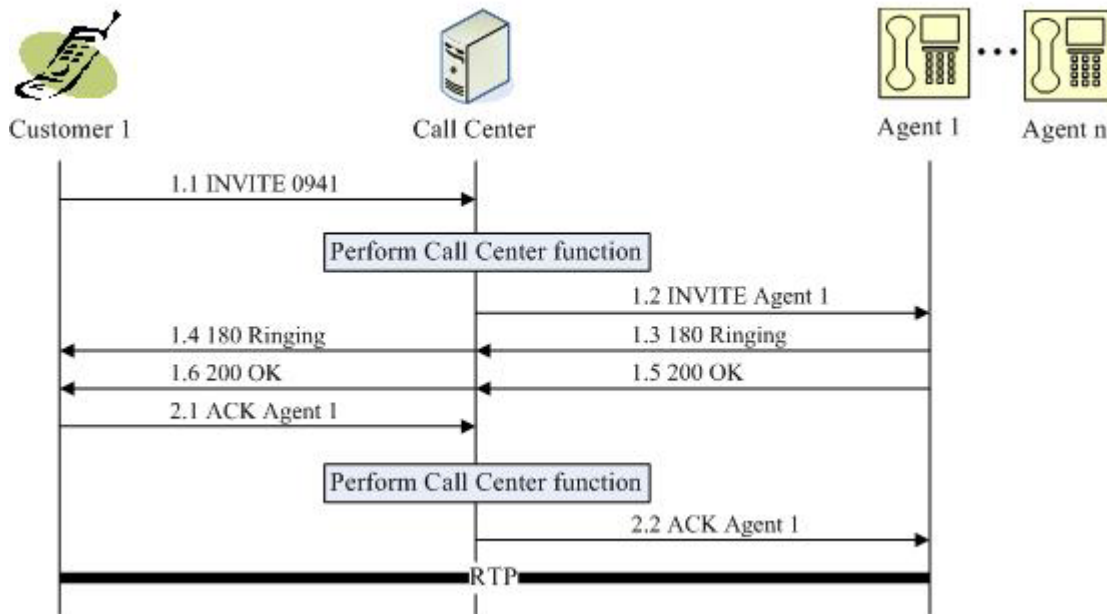


Figure 4.6. Normal call setup message flow.

Step 1. When Customer 1 dials the call center’s identity “0941”, the call center passes the SIP INVITE message to the agent dispatcher (see Steps 1–4 in Figure 4.7). The `select_agent()` function (Figure 4.7 (5)) selects an agent with the IDLE state (i.e., Agent 1) to serve this inbound call and rewrite the SIP INVITE message’s target URI to Agent 1’s. The Call-ID is then stored in the ACD database. The `set_agent_state()` function (Figure 4.7 (7)) sets Agent 1’s status to PENDING. The SIP INVITE message is then forwarded to Agent 1. When Agent 1 accepts the call, it sends back a SIP 200 OK message.

Step 2. Customer 1 sends the SIP ACK message to the call center after a final response 200 OK message is received. The `set_agent_state()` function (Figure 4.7 (11)) sets Agent 1’s status to TALKING. If the request URI is call center’s

identity (Figure 4.7 (12)), the agent dispatcher executes the `rewrite_request_uri()` function (Figure 4.7 (13)) to forward this message to the corresponding agent. Then the media connection between Customer 1 and Agent 1 is established, the media data is transferred by RTP.

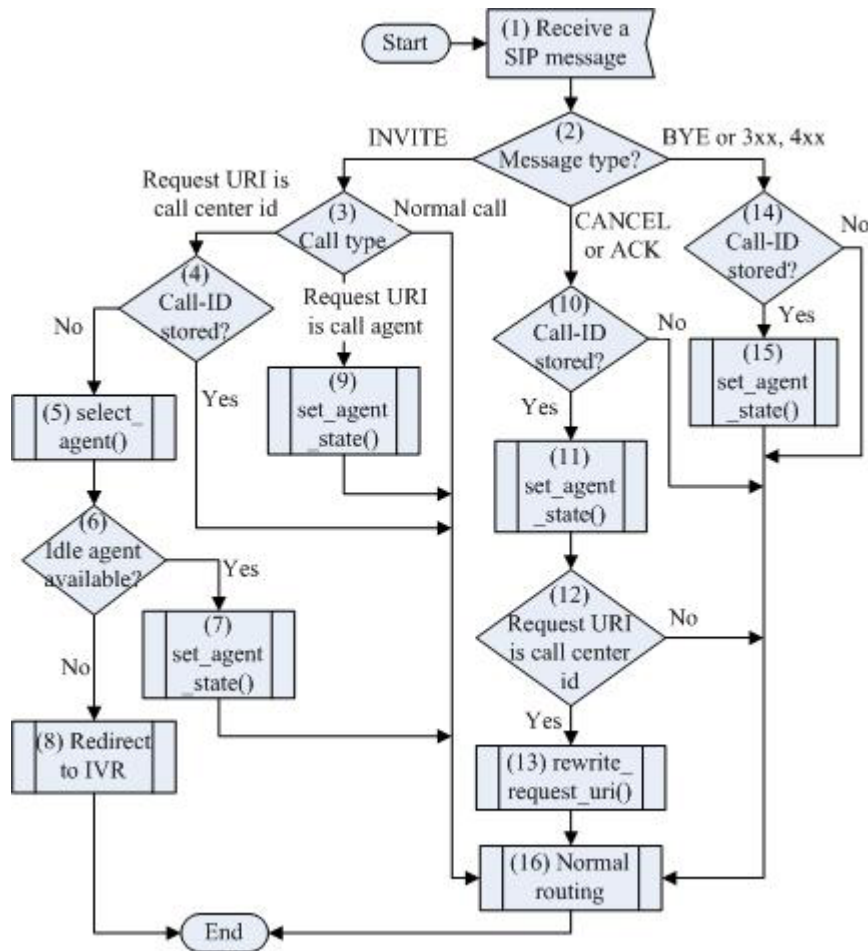


Figure 4.7. The flow diagram for the agent dispatcher.

If all agents are busy, the call center redirects inbound calls to the IVR to announce the predicted waiting time. The call setup procedure when all agents are busy is illustrated in Figure 4.8, which consists of the following steps:

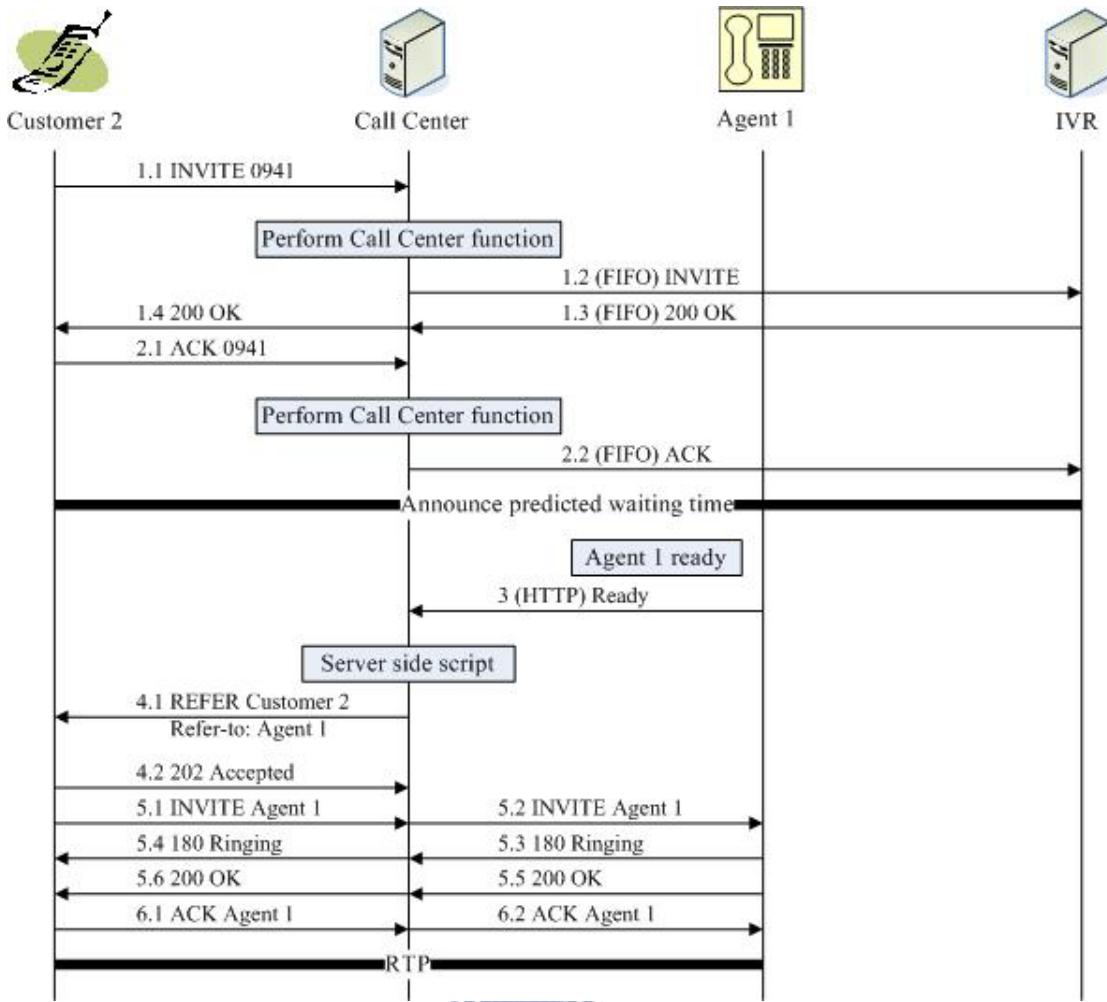


Figure 4.8. Message flow when all agents are busy.

Step 1. Customer 2 attempts to access the call center when all agents are busy. The `select_agent()` function tries to select an idle agent and cannot find any. It then calculates the predicted waiting time (Figure 4.7 (5)). The agent dispatcher notifies the predicted waiting time to the IVR (Figure 4.7 (8)). The SIP INVITE message is then redirected to the IVR.

Step 2. Customer 2 sends the SIP ACK message to the call center upon receipt of the final response 200 OK message. Since the Call-ID of this message is not stored (Figure 4.7 (10)), the agent dispatcher normally routes the SIP ACK message to the IVR (Figure 4.7 (16)). The media connection between Customer 2 and the

IVR is established. The IVR announces the predicted waiting time to Customer 2.

Step 3. Agent 1 becomes available and notifies the call center that it is ready to serve.

Step 4. The call center selects the first request in the queue and sends a SIP REFER message to notify Customer 2 that the call is connected to Agent 1.

Step 5. Base on the SIP standard [6], after the reception of the SIP REFER message, Customer 2 has to send a SIP INVITE message with the URI of Agent 1 which is indicated in the Refer-to header of the SIP REFER message. The `set_agent_state()` function (Figure 4.7 (9)) sets Agent 1's status to PENDING. The SIP INVITE message is then forwarded to Agent 1.

Step 6. Customer 2 sends the SIP ACK message to the call center after a final response 200 OK message is received. The `set_agent_state()` function (Figure 4.7 (11)) sets Agent 1's status to TALKING. After the SIP ACK message is forwarded to Agent 1, the media connection between Customer 2 and Agent 1 is established.



4.4 The Waiting Time Prediction Algorithm

The waiting time prediction algorithm executed in the `select_agent()` function is described as follows. Let n be the number of agents, λ be the call arrival rate, and μ be the service rate. Let $k[i]$ be the queue length observed by the i th customer who is queued at his/her arrival (i.e., all agents are busy when the customer arrives), $Wr[i]$ be the actual waiting time for customer i , and $Wp[i]$ be the predicted waiting time. Based on the M/M/n model [41, 42], Whitt [43, 44] predicts the waiting time for customer i as

$$Wp[i] = \frac{k[i] + 1}{n\mu} \quad (4.1)$$

It is important to evaluate the accuracy of (1), which was not mentioned in [43], and is conducted in this section as follows. Define δ as the accuracy measurement expressed as:

$$\delta = \frac{\sum_{i=1}^M |Wp[i] - Wr[i]|}{\sum_{i=1}^M Wr[i]}, \text{ where } M \text{ is the number of customers who are queued in}$$

the observation period. It is apparent that the smaller the δ value, the better the prediction. Based on our “customer experience”, a waiting customer is frustrated if the actual waiting time is longer than the predicted waiting time [45]. Therefore, another accuracy measurement θ is the probability that the actual waiting time is longer than the predicted waiting time. That is,

$$\begin{aligned} \theta &= \Pr(Wp[i] < Wr[i]) \\ &= \frac{\text{the number of customers } i \text{ such that } Wp[i] < Wr[i]}{M} \end{aligned} \quad (4.2)$$

Equation (2) implies that the smaller the θ value, the better the prediction. Note that θ and δ may be conflicting. For example, if we announce a very large $Wp[i]$ value for all customers i , then $\theta = 0$ but $\delta = 1$. Therefore, it is important to predict the waiting time “wisely” such that both θ and δ are sufficiently small. A simple approach to improve the θ measurement is to introduce a factor $\alpha > 1$ in (1). That is, (1) is re-written as

$$Wp[i] = \frac{\alpha(k[i] + 1)}{n\mu}, \quad \alpha \geq 1$$

4.4.1 Enhanced Whitt’s Algorithm

We develop an algorithm to dynamically select α that limits θ to a targeted number β (say 0.3) while keeping δ at a sufficiently small value. The steps are described as follows.

Step 0. Select an initial α value, predefined β value, and let $i = 0$.

Step 1. For the next call arrival, if there are idle agents, then route the call to an idle agent. Repeat this step. Otherwise, increment i by 1, and go to Step 2.

Step 2. Calculate $Wp[i] = \frac{\alpha(k[i]+1)}{n\mu}$. If i is larger than a threshold (say 1,000), go to Step 3. Otherwise, go to Step 1.

Step 3. Calculate θ based on (4.2), and reset $i = 0$.

Step 4. If $\theta > \beta$ (the predicted waiting time is too small), then $\alpha = \alpha + \gamma$. Otherwise $\alpha = \alpha - \gamma$ (γ is a positive value, say 0.01).

Step 5. Go to Step 1.

When an agent becomes available and the queue is not empty, the queue manager will route the first queued customer to the agent.

The above algorithm does not capture the behavior of the customers being served by the agents, and is thus further modified as follows. Let $w_l[i]$ be the *elapsed service time* that an agent l has already spent on the served customer when the i th customer arrives, and the *residual service time* be the period left before agent l completes the served call. Note that the $Wp[i]$ equation in Step 2 simply assumes that when a new customer arrives, the residual service times of the agents are the same as the service times of the waiting customers. This assumption is not true if the service time distribution is not Exponential. To accommodate the effect of the "residual service times", we modify the above algorithm by considering a set of α values with the following intuition. If two customer arrivals "see" the same amount of net elapsed service times at the call agents, then these two customers will have the same α value in computing the $Wp[i]$ equation. In other words, we classify the customer arrivals based on the "net elapsed service times" spent at the agents. For the i th customer, the net elapsed service time at his/her arrival is

$$T[i] = w_1[i] + w_2[i] + \dots + w_n[i]$$

Consider the following heuristics to map $T[i]$ to an integer j_i .

$$j_i = \left\lfloor \frac{T[i]}{n\mu \cdot Vs^{1/2}} \right\rfloor \quad (4.3)$$

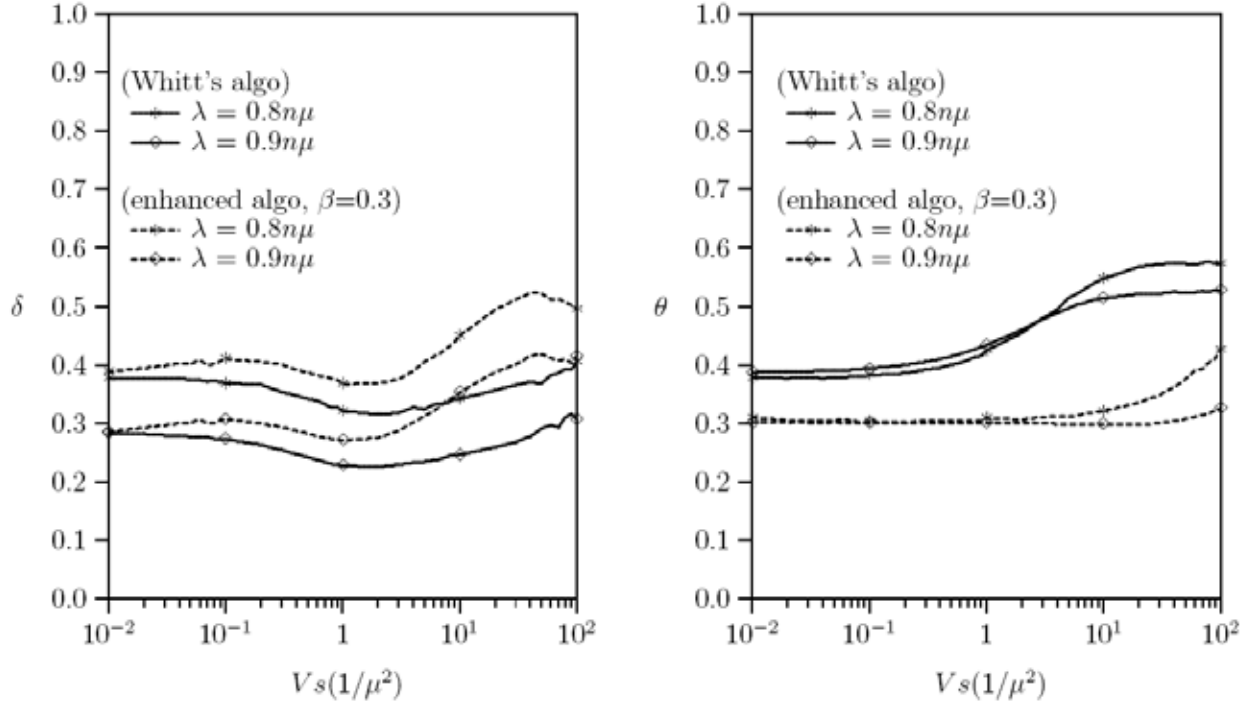
where the mean $1/\mu$ and the variance Vs are computed through the service time samples. Note that there are many mapping alternatives, which show similar results and will not be discussed here. Based on (4.3) we modify the above algorithm by considering two arrays for α and θ , respectively. Specifically, $\alpha[j_i] = \alpha[j_m]$ if $j_i = j_m$ at Step 2; that is, $Wp[i]$ is computed as

$$Wp[i] = \frac{\alpha[j_i](k[i] + 1)}{n\mu}$$

At Step 3, $\theta[j_i]$ is calculated dynamically for every 1000 customer arrivals " m " such that $j_i = j_m$, and $\alpha[j_i]$ is computed based on $\theta[j_i]$ at Step 4.

4.4.2 Performance Evaluation

We develop a discrete event simulation model to investigate δ and θ of waiting prediction algorithms. Note that the θ performance is a weighted average of $\theta[j_i]$. In the simulation experiments, the customer arrivals are a Poisson stream with rate λ , and the service times have a Gamma distribution with mean $1/\mu$ and variance Vs . The Gamma distribution is selected because it has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions (see Lemma 3.9 in [46]). Each experiment simulates more than 10 million customer arrivals to ensure that the results are stable. Figure 9 shows the δ and θ performances, where $n=80$ and $\mu=0.05s^{-1}$. Based on the experiments, we have the following observations.



(a) The δ performance

(b) The θ performance

Figure 4.9. The δ and θ performances ($n=80, \mu=0.05/\text{sec}, \gamma=0.01$).

- **Effect of Customer Arrival Rate λ :** Figure 4.9 (a) shows that δ increases as λ decreases. In other words, the predicted algorithms are more accurate as the customer arrival rate increases.
- **Effect of Service Time's Variance V_s :** In Figure 4.9 (a), the smallest δ is observed when $V_s = \mu^{-2}$ (when the service times have an exponential distribution) because $Wp[i]$ is derived based on exponential service times. As V_s increases or decreases from μ^{-2} , δ increases.
- **Effect of α :** The δ values for the enhanced algorithm is larger than that for the Whitt's algorithm as expected. This effect becomes significant where V_s is large. As shown in Figure 9 (b), the θ performance implies that larger δ values of the enhanced algorithm are contributed by smaller actual waiting times. That is, a waiting customer is likely to be answered earlier than he/she expects. Such "inaccuracy" is reasonably acceptable to the customers, according to mobile

operators [45]. Also, when V_s is small (which is likely to be true for routine call agent services), the enhanced algorithm can significantly outperform the Whitt's algorithm for the θ performance (28.9%) without significantly degrading the δ performance (0.7%).

- **Effect of targeted probability β :** Figure 4.9 (b) shows that when β is set to 0.3, the enhanced algorithm can maintain θ that is close to the targeted β value. For instance, when $\lambda = 0.9n\mu$, $V_s = \mu^{-2}$, θ is 0.434 in Whitt's algorithm while θ is 0.3 in the enhanced algorithm.
- **Effect of impatient customers:** Figure 4.10 shows the δ and θ performances by assuming that the waiting customers may be impatient and leave the system. A customer might decide to abandon the call request immediately after hearing the predicted waiting time announcement or after he/she has waited for a period. When customer i hears the predicted waiting time announcement $Wp[i]$, he/she compares $Wp[i]$ with the amount of time he/she is willing to wait, which is denoted as $DropTime[i]$. If $Wp[i] > DropTime[i]$, the customer abandons the call immediately. Otherwise, the customer chooses to wait. If the actual waiting time $Wr[i]$ is longer than $Wp[i]$, the waiting customer may be impatient and leaves the system. Our experiments assume that customer i gives up waiting after a period of time $PatientTime[i]$, where $PatientTime[i] = Wp[i] + x[i]$, and $x[i]$ is a gamma-distributed random variable with mean $Mx[i]=0.3*Wp[i]$ and variance $Vx[i]$. $DropTime[i]$ is a gamma-distributed random variable with mean μ^{-1} seconds and variance $0.5*\mu^{-2}$. Figure 10 shows the δ and θ performances where $\lambda=0.8n\mu$ and $\beta=0.1$. It can be seen that the benefit of the enhanced algorithm in θ performance becomes insignificant.

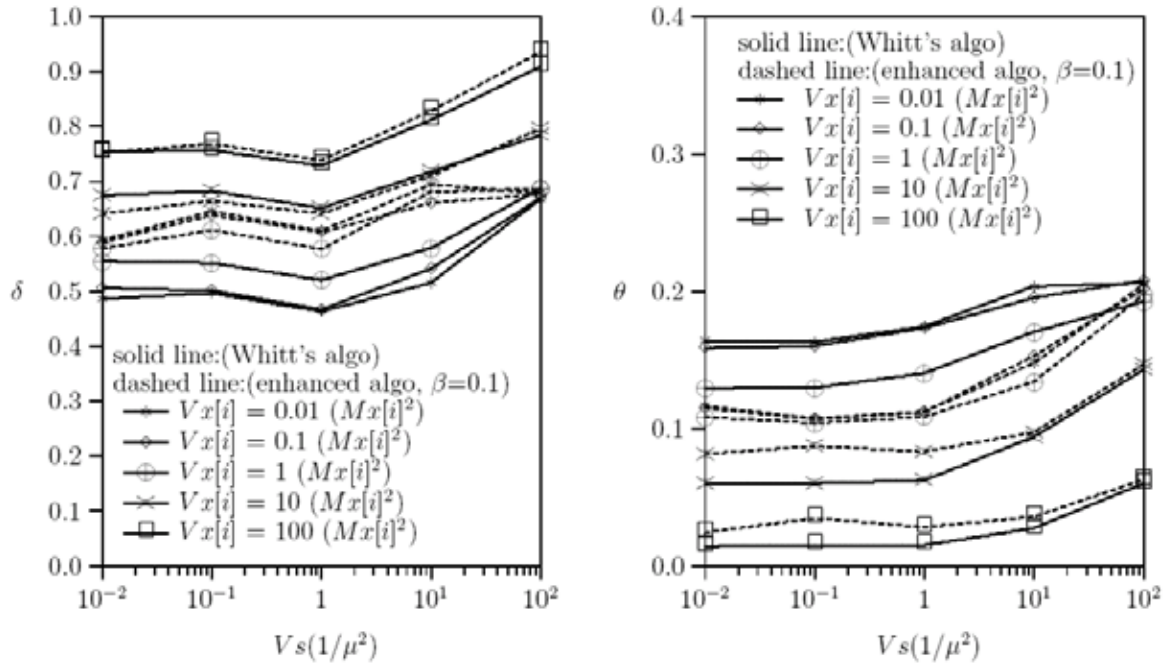


Figure 4.10. The δ and θ performances when customers may be impatient.

4.5 Callback Mechanism

When all agents are busy in a call center, customers might decide to leave the system after they heard the predicted waiting time announcement or after they have waited for a period of time. When the agents are free, they may still serve those customers through the callback mechanism. However, callback does not work when a customer resides in a *Private Telecommunications Network* (PTN).

A PTN is a telecommunications network that has its own number plan other than the public E.164 numbering used in the Public Switched Telephone Network (PSTN) [47]. Examples of PTN are enterprise telephone systems and VoIP networks without being assigned PSTN E.164 numbers. Figure 4.11 illustrates an abstract PTN architecture that interconnects to the PSTN. The core component in this architecture is the *Private Branch Exchange* (PBX) in a telephony-based PTN or the *VoIP PSTN Gateway* (VPG) in an Internet-based PTN. Without loss of generality, we consider the Internet-based PTN. An example of the Internet-based PTN is Skype [48].

Suppose that there are N users in the Internet-based PTN. These users connect to the VPG through Internet or enterprise Intranet. Note that there may be more than one VPGs installed in the PTN. A VPG connects to a switch in the PSTN with n leased lines (where $n \leq N$). In a hybrid architecture, the VPG may indirectly connect to the PSTN through a PBX.

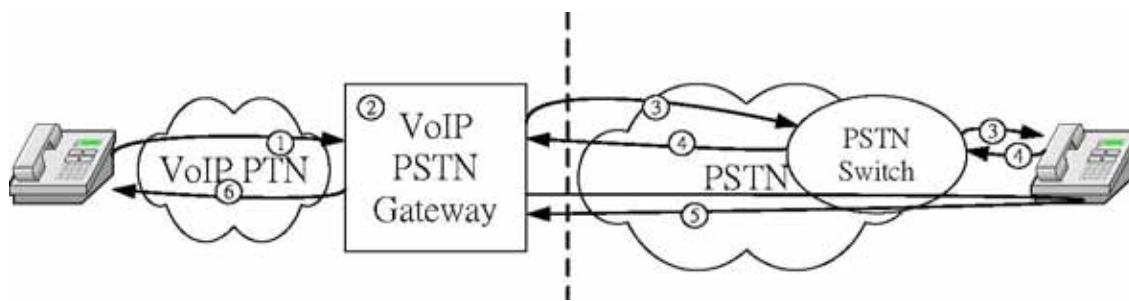


Figure 4.11: Private Telecommunications Network Architecture.

When a PTN user attempts to make a call to a PSTN user (this procedure is referred to as *PTN call origination*), the call is first set up to the VPG (Figure 4.11 (1)) using VoIP signaling such as SIP, H.323 [4, 49], or *Media Gateway Control Protocol* (MGCP) [50, 51]. The VPG then chooses an available leased line (among the n -leased-line pool; Figure 4.11 (2)), and connects the call to the PSTN switch (Figure 4.11 (3)) using *Signaling System Number 7* (SS7) [25].

When a PSTN user (an agent of the call center in our example) attempts to call a PTN user (this procedure is referred to as the *PTN call termination*), he/she first dials the telephone number of the VPG (i.e., the number is one of the n leased lines owned by the VPG). The PSTN switch connecting the caller (a call agent) will set up the call to the VPG using SS7 (Figure 4.11 (4)). Then the VPG will request the PSTN caller (e.g., through the IVR system) to input the extension number of the called PTN user (Figure 4.11 (5)). After the PSTN caller has input the extension number (e.g., through DTMF dialing [39]), the VPG sets up the call to the called PTN user through VoIP protocol (Figure 4.11 (6)). In this call termination procedure, a PTN user is accessed indirectly through two-stage dialing. That is, the call agent must first dials the number

of the VPG of the PTN, and then dials the extension number of the called PTN user.

In a call setup, the SS7 Initial Address Message (IAM) [25] delivers the caller's telephone number (i.e., the *caller ID*) to the called party. The caller ID will be automatically stored in the called party's telephone device (e.g., a mobile phone with address book). The called party can then call back without dialing the caller's telephone number.

A limitation of PTN is that from the viewpoint of the PSTN, a PTN user does not have a caller ID. When a PTN user reaches a PSTN user (a call center in our example), the PTN's identity cannot be carried by the SS7 IAM. Instead, the PBX or VPG's telephone number is used as the caller ID. Consequently, the call agent cannot utilize the callback service to reach the PTN user later. This limitation may cause frustration to the agent in the call center, especially if the call agent is not available when the PTN user calls the call center. The call agent will not be able to call back based on the caller ID, and he/she does not even know who made the call previously. In this case, important calls may be lost.

To resolve the above issue, we propose a mechanism to allow a call center outside a PTN to call back a user within the PTN.

4.5.1 The PTN Callback Mechanism

The PTN callback mechanism is established by introducing a *callback table* in the VPG. An entry of the table consists of three fields. The first field records the PTN caller's identity (including the user's name, phone number, SIP URI, or any ID representation). The second field records the PSTN called party's (the call center) telephone number. The third field records the time when the entry is created. In PTN call origination, the PTN caller's ID information and the PSTN call center's telephone number are stored in the callback table at Step (2) in Figure 4.11. Then Step (3) is executed to set up the call. Note that before Step (3) is actually carried out,

or when the call is terminated (especially when all agents in the call center are busy and the call does not get through), the calling PTN user may specify if callback mechanism should be triggered for this call. If not, the callback entry will not be created, and the VPG may request the PSTN switch to disable the caller ID feature when the call is set up to the call center.

When the call center calls back using the standard PSTN callback procedure, the call setup follows the same call termination flow described in Figure 4.11 except that at Step (5), the VPG's IVR first asks the call agent if this call is a callback. (Alternatively, the VPG may check the callback table first. If no entry is found, the IVR will not ask the call agent if this call is a callback.) If the call agent's answer is negative, the VPG executes the normal call termination procedure. If the call agent's answer is positive, the VPG will check the callback table by using the call center's telephone number as the index to retrieve the PTN user's identity. If no entry is found, the call is terminated. If exact one entry is found, then the call is set up to the target PTN user directly. If two or more entries are found, the IVR announces all PTN user names of the retrieved entries (e.g., *Jason Lin*, *MF Chang*, *Honda Hsu* and so on). The call agent then identifies the person to be called back, and the VPG sets up the call to the corresponding address following Step (6) in Figure 4.11.

After a callback is established, depending on the table management policy, the corresponding entry in the callback table may or may not be removed.

4.5.2 Discussions

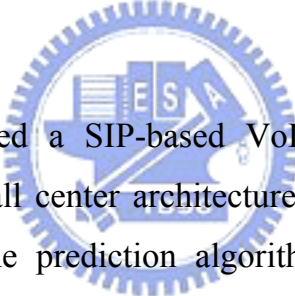
The callback mechanism described in the section 4.5.1 utilizes IVR for callback setup. The IVR involvement can be eliminated with the following arrangements at the VPG.

- The n leased lines to the PSTN are grouped into two categories: *normal leased line set* and *callback leased line set*. The phone numbers in the normal set are announced to the public while the phone numbers in the callback set are not

visible to the PSTN users for normal PTN call termination. In call origination, if the calling PTN user triggers callback at Step (2), then the VPG use a leased line in the callback set to connect the call. If not, the call is set up using a leased line from the normal set.

- In call termination, if the call agent dials a phone number in the normal set, then the call is set up following the normal call termination procedure. On the other hand, if a number in the callback set is dialed, the VPG considers it as a callback. The VPG retrieves the callback table to identify the target PTN user, and sets up the call to that user. If two or more entries are found, the VPG may choose to connect the PTN user of the entry with the latest timestamp, and may also inform the call agent that there are other callback requests.

4.6 Conclusions



In this chapter, we proposed a SIP-based VoIP call center with waiting time prediction. The SIP-based call center architecture and detailed message flows were elaborated. The waiting time prediction algorithms are more accurate when the customer arrival rate is larger. Furthermore, the enhanced algorithm can effectively control (and limit) the probability θ that a customer will be queued longer than the predicted waiting time. We proposed a callback mechanism that allows a call center agent to call back a customer in the PTN. The callback table approach is utilized as a plug-in solution that only needs to insert the callback table module into the VoIP PSTN Gateway. Our solution does not affect the existing call setup message flow. As a final remark, the callback mechanism proposed in this paper is pending USA and ROC patents.

Chapter 5

Conclusions and Future Work

In this dissertation, we investigated three design issues on the call control of IP multimedia services. This chapter summarizes our study and contributions, and briefly discusses the future directions.

5.1 Summary

In this dissertation, we discussed three IP multimedia services design issues. In Chapter 2, an integrated call agent of the converged VoIP network was proposed. We presented a simple, flexible framework for the interworking functions of VoIP protocols based on IN half-call BCSM. In addition, we have implemented the basic gateway components, O_BCSMs and T_BCSMs, for SIP, H.323, and MGCP. The caller FSM of one VoIP protocol can interact with the callee FSM of any VoIP protocol. The development effort of the interworking function is minimized since only two half-call FSMs for each VoIP protocol are needed and they can be developed independently. A converged VoIP network can be managed by a group of coordinating ICAs such that two user devices managed by different ICAs can communicate. The limitation of this approach is that the interworking function for a P2P VoIP system and a client-server one (such as SIP) needs to be investigated.

In Chapter 3, we proposed an one-pass authentication, in which the redundant steps in the GPRS authentication and IMS authentication procedures are removed. Compared with the eight-step two-pass authentication, the four-step one-pass authentication saves two to four SIP/Cx message exchanges among the MS, the SGSN, the CSCF, and the HSS/AuC. The one-pass authentication can save up to 50% of the network traffic generated by the IMS registration, and saves 50% of the storage for buffering the authentication vectors. We also formally proved that the IMS user is correctly authenticated in the one-pass authentication. The limitation of this approach is that a SIP ALG is required in the one-pass procedure. Since IMS is based on SIP, a SIP ALG is required for other purposes. Therefore, the one-pass procedure will incur little extra cost for implementing SIP ALG.

In chapter 4, we described the design and implementation of a SIP-based VoIP call center with waiting time prediction. The SIP-based plug-in modular call center architecture and detailed message flows were elaborated. We proposed two output measures and developed a discrete event simulation model to investigate the performance of the waiting time prediction algorithm for the call center. The waiting time prediction algorithms are more accurate when the customer arrival rate is larger. Furthermore, the enhanced algorithm can effectively control (and limit) the probability θ that a customer will be queued longer than the predicted waiting time. The limitation of this approach is that the prediction algorithms are less accurate as the customer arrival rate decreases or the service time's variance grows too large.

5.2 Future Works

Based on the research results in this dissertation, the following design issues on the IP multimedia services network can be investigated further.

Interworking with P2P VoIP clients: Recently, some successful cases of P2P VoIP communications such as Skype have become very popular. The interworking function for a P2P VoIP system and a client-server one (such as SIP) is an important issue that needs to be investigated. Furthermore, many instant message applications such as Yahoo! Messenger, MSN Messenger and GTalk from Google have provided Internet voice and video services. The routing, naming and interworking issues in a converged network which integrates the above applications can be further discussed.

Integrated WiFi and application level authentication: While WiFi was employed as one of the access network of UMTS, the one-pass authentication may be integrated into the UMTS and WiFi interworking network. In addition, the IMS application service platform provides a flexible environment for the third party to run their own applications. Before the third party provides the services, the MS has been authenticated at the GPRS level, the IMS level, and maybe the application service platform. We can further discuss the integration of the authentication procedures in the GPRS, IMS, and application platform.

Call center emulator with prediction algorithm: In the call center, it is very useful to monitor the performance of the prediction algorithm. With the logs and statistics of the performance we could fine tune the prediction algorithm. The emulation of the call center could also provide useful information while managing the call center.

Reference

- [1] ITU, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-CELP)," *Recommendation G.729*, ITU-T.1, Geneva, 1996.
- [2] S. Casner, R. Frederick, V. Jacobson, and H. Schulzrinne, "RTP: A Transport Protocol for Real-Time Applications," *IETF Request for comments 1889*, IETF, 1996.
- [3] H. Schulzrinne, "RTP Profile for Audio and Video Conferences with Minimal Control," *IETF Request for comments 1890*, IETF, 1996.
- [4] ITU, "Packet-Based Multimedia Communication Systems," *ITU-T Recommendation H.323*, ITU, 1998.
- [5] ITU, "Call signaling over UDP", *ITU-T Recommendation H.323 Annex E*, ITU, 1998.
- [6] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," *IETF Request for Comments 3261*, IETF, 2002.
- [7] M. Arango, A. Dugan, I. Elliott, C. Huitema, and S. Pickett, "Media Gateway Control Protocol (MGCP) Version 1.0," *IETF Request for Comments 2705*, IETF, 1999.
- [8] M. Handley and V. Jacobson, "SDP: Session Description Protocol," *IETF Request for Comments 2327*, IETF, 1998.
- [9] C. Groves and M. Pantaleo, "The Megaco/H.248v2 Gateway Control Protocol, version 2," *IETF Internet draft draft-ietf-megaco-h248v2-03.txt*, IETF, 2002.
- [10] EURESCOM: 'Providing IN functionality for H.323 telephony calls,' Project report P916, Research and Strategic Studies in Telecommunications, European Institute, 1999.
- [11] K.V. Vemuri, "SPHINX: A Study in Convergent Telephony," *IP Telecom Services Workshop (IPTS2000)*, Georgia, USA, 2000, pp. 9-18.
- [12] K. Singh and H. Schulzrinne, "Interworking Between SIP/SDP and H.323," *1st IP-Telephony Workshop (IPTEL2000)*, Berlin, 2000, pp. 75-92.
- [13] H. Agrawal, V. Palawat, and R. Roy, "SIP-H.323 Interworking," *IETF Internet draft draft-agrawal-sip-h323-interworking-reqs-02.txt*, IETF, 2001.

- [14] V. Gurbani and V. Rastogi: 'Accessing IN Services from SIP networks,' IETF Internet draft draft-gurbani-iptel-sip-to-in-05.txt, 2001.
- [15] F. Haernes, "SIP-IN Interworking Protocol Architecture and Procedures," *IETF Internet draft draft-haerens-sip-in-00.txt*, IETF, 2001.
- [16] R. Ackermann, V. Darlagiannis, M. Goertz, M. Karsten, and R. Steinmetz, "An Open Source H.323 / SIP Gateway as Basis for Supplementary Service Interworking," *2nd IP Telephony Workshop (IPTEL2001)*, New York, USA, 2001, pp. 1-7.
- [17] W. Jiang, J. Lennox, S. Narayanan, H. Schulzrinne, and K. Singh, "CINEMA: Columbia InterNet Extensible Multimedia Architecture," Technical Report CUCS-011-02, Department of Computer Science, University of Columbia, 2002.
- [18] U. Black, *The intelligent network: Customizing telecommunication networks and services*. Prentice Hall, New Jersey, 1998.
- [19] 3GPP, 3rd generation partnership project; technical specification group services and systems aspects; 3G security; security architecture, Tech. Spec. 3G TS 33.102 V3.7.0 (2000–12), 2000.
- [20] 3GPP, 3rd generation partnership project; Technical specification group services and systems aspects; General packet radio service (GPRS); Service description; Stage 2, Tech. Spec. 3G TS 23.060 version 4.1.0 (2001–06), 2001.
- [21] 3GPP, 3rd generation partnership project; Technical specification group services and systems aspects; IP Multimedia subsystem stage 2, Tech. Spec. 3G TS 23.228 version 6.2.0 (2003–06), 2003.
- [22] 3GPP, 3rd generation partnership project; Technical specification group core network; Signaling flows for the IP multimedia call control based on SIP and SDP; Stage 3, version 5.5.0 (2003–06). 3GPP TS 24.228, 2003.
- [23] Y.-B. Lin and Y.-K. Chen, "Reducing authentication signaling traffic in third generation mobile network," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 493–501, 2003.
- [24] 3GPP, 3rd generation partnership project; Technical specification group services and systems aspects; 3G security; Access security for IP-based services, Tech. Spec. 3G TS

- 33.203 V5.5.0 (2003–03), 2003.
- [25] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York: Wiley, 2001.
- [26] V. W.-S. Feng, L.-Y. Wu, Y.-B. Lin, and W. E. Chen, “WGSN: WLANbased GPRS environment support node with push mechanism,” *Comput. J.*, vol. 47, no. 4, pp. 405–417, 2004.
- [27] Y.-B. Lin, Y.-R. Huang, Y.-K. Chen, and I. Chlamtac, “Mobility management: From GPRS to UMTS,” *Wireless Commun. Mobile Comput.*, vol. 1, no. 4, pp. 339–360, 2001.
- [28] Y.-B. Lin, Y.-R. Hanug, A.-C. Pang, and I. Chlamtac, “All-IP approach for UMTS third generation mobile networks,” *IEEE Netw.*, vol. 16, no. 5, pp. 8–19, 2002.
- [29] 3GPP, 3rd generation partnership project; Technical specification core network; Cx and Dx interfaces based on the diameter protocol; Protocol details, Tech. Spec. 3G TS 29.229 V5.3.0 (2003–03), 2003.
- [30] 3GPP, 3rd generation partnership project; Technical specification core network; IP multimedia subsystem Cx and Dx interfaces; Signaling flows and message contents (Release 5), Tech. Spec. 3G TS 29.228 V5.4.0 (2003–06), 2003.
- [31] W. E. Chen, Q. Wu, A.-C. Pang, and Y.-B. Lin, “Design of SIP application level gateway for UMTS,” in *Design and Analysis of Wireless Networks*, Y. Pan and Y. Xiao, Eds. Commack, NY: Nova, 2004.
- [32] Y.-B. Lin, P.-J. Lee, and I. Chlamtac, “Dynamic periodic location area update in mobile networks,” *IEEE Trans. Veh. Technol.*, vol. 51, no. 6, pp. 1494–1501, 2002.
- [33] Y.-B. Lin, M.-F. Chen, and H. C.-H. Rao, “Potential fraudulent usage in mobile telecommunications networks,” *IEEE Trans. Mobile Comput.*, vol. 1, no. 2, 2002.
- [34] A. Szlam, and K. Thatcher, *Predictive Dialing Fundamentals*. Melita International Corporation, 1996.
- [35] A.-C. Pang, C.-H. Liu, S.-P. Liu, and H.-N. Hung, “A Study on SIP Session Timer for

- Wireless VoIP,” IEEE Wireless Communications & Networking Conference (WCNC), New Orleans, USA, 2005.
- [36] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” IETF, RFC 3550, July 2003.
- [37] A. Miloslavski, V. Antonov, etc, “Third-party Call Control in VoIP Networks for Call Center Applications,” *IEEE Intelligent Network Workshop*, 2001.
- [38] S. R. Ahuja, J. R. Ensor, “VoIP: What is it Good for?” *ACM Queue*, 2004.
- [39] Y.-B. Lin, *Introduction to Telephony*. Vekeg, 1997.
- [40] <http://www.iptel.org>, 2005.
- [41] L. Kleinrock, *Queueing Systems*, Volume 1: Theory. John Wiley & Sons, 1975.
- [42] L. Kleinrock, *Queueing Systems*, Volume 2: Computer Applications. John Wiley & Sons, 1976.
- [43] W. Whitt, “Improving Service by Informing Customers About Anticipated Delays,” *Management Science*, Vol. 45, No. 2, February 1999.
- [44] W. Whitt, “Predicting queueing delays,” *Management Science*, Vol. 45, No. 6, 1999.
- [45] FarEastone, Private communications, 2006.
- [46] F.P. Kelly, *Reversibility and stochastic Networks*. John Wiley & Sons, 1979.
- [47] CCITT, “Numbering Plan for the ISDN Era,” Technical Report Recommendation E.164 (COM II-45-E), ITU-T, 1991.
- [48] Skype. <http://www.skype.com>, 2002.
- [49] C.-H. Rao, Y.-B. Lin, and S.-L. Chou, “iGSM: VoIP Service for Mobile Networks,” *IEEE Communications Magazine*, 4(38):62–69, 2000.
- [50] F. Andreassen, and B. Foster, “Media Gateway Control Protocol (MGCP) Version 1.0,” IETF, RFC3435, 2003.
- [51] M.-F. Chang, Y.-B. Lin, and A.-C. Pang, “vGPRS: A Mechanism for Voice over GPRS,” *ACM Wireless Networks*, 2001.

- [52] “New WCDMA, HSDPA and EDGE Surveys by GSA Confirm Accelerating Mobile Broadband Growth,” *The Global mobile Suppliers Association*, 2006.
- [53] 3GPP, <http://www.3gpp.org>, 2006.
- [54] 3GPP2, <http://www.3gpp2.org>, 2006.
- [55] 3GPP, 3rd generation partnership project; Technical specification group services and system aspects; 3GPP enablers for Open Mobile Alliance (OMA); Push-to-talk over Cellular (PoC) services; Stage 2 (Release 6), Tech. Spec. 3G TR 23.979 V6.2.0 (2005–06), 2005.
- [56] Y.-B. Lin and A.-C. Pang, *Wireless and Mobile All-IP Networks*. Wiley, 2005.
- [57] G. Horn, D. Kröselberg, and K. Müller, “Security for IP multimedia services in the 3GPP third generation mobile system,” *Internet Research*, Vol. 13, No. 2, 2003.



Curriculum Vitae

Hsu, Meng-Ta was born in Changhua, Taiwan, R.O.C., in 1975. He received his B.S. degree in Computer Science & Information Engineering (CSIE) from National Taiwan University (NTU) in 2000. He is currently a Ph.D. candidate in CSIE, National Chiao-Tung University (NCTU). His research interests include Internet communications, Voice over IP (VoIP) networks, intelligent transportation systems, mobile computing and performance modeling.



Publication List

- **Journal Publications**

1. Y.-B. Lin, M.-F. Chang, Meng-Ta Hsu, and L.-Y. Wu, “One-Pass GPRS and IMS Authentication Procedure for UMTS,” *IEEE Journal on Selected Areas in Communications*, 23(6): 1233-1239, June 2005
2. H.-H. Chang, Meng-Ta Hsu, M.-F. Chang, “An Integrated Call Agent of the Converged VoIP Network,” accepted and to appear in *Journal of Information Science and Engineering*
3. Meng-Ta Hsu, Yi-Bing Lin, Bo Li, and Ming-Feng Chang, “A SIP-based Call Center with Waiting Time Prediction,” accepted and to appear in *Journal of Internet Technology*

