國立交通大學

資訊工程學系

博士論文

提升固定式與行動通訊整合環境之號碼可攜服務效能

Enhancing The Efficiency Of Number Portability Service On

Fixed-Mobile Convergence Environment

研 究 生 ： 鄭靜紋

指導教授 ： 鍾乾癸 博士

中華民國九十六年七月

提升固定式與行動通訊整合環境
之號碼可攜服務效能

Enhancing The Efficiency of Number Portability Service on Fixed-Mobile Convergence Environment

研究生: **鄭靜紋** Student: Ching-Wen Cheng

指導教授 : **鍾乾癸 博士** Advisor: Dr. Chyan-Goei Chung

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

July 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年七月

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
___________組， __95__ 學年度第_2__學期取得博士學位之論文。
論文題目：提升固定式與行動通訊整合環境之號碼可攜服務效能
指導教授：鍾乾癸

■ 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	■ 中華民國 96 年 7 月 31 日公開
校外網際網路	■ 中華民國 96 年 7 月 31 日公開

授 權 人：鄭靜紋

親筆簽名：______________________

中華民國　96　年　7　月　31　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
____________組， 95　 學年度第_2__學期取得博士學位之論文。
論文題目：提升固定式與行動通訊整合環境之號碼可攜服務效能
指導教授：鍾乾癸

■ 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：鄭靜紋

親筆簽名：______________________

中華民國　96　年　7　月　31　日

國家圖書館博碩士論文電子檔案上網授權書

ID:GT008617815

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 _________ 組 _95_學年度第_2_學期取得博士學位之論文。

論文題目：提升固定式與行動通訊整合環境之號碼可攜服務效能

指導教授：鍾乾癸

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：鄭靜紋

親筆簽名：_______________

民國 96 年 7 月 31 日

提升固定式與行動通訊整合環境之號碼可攜服務效能

學生：鄭靜紋　　　　　　　　　　　　　　指導教授：鍾乾癸博士

國立交通大學資訊工程系博士班

摘要

在通訊環境整合之前，固接式電話(fixed-lined)與行動電話(mobile)通常各別維運，通訊服務也各自獨立。固定式與行動通訊整合環境（Fixed-mobile convergence, FMC）提供一個可供固接式與行動電話業者互通並分享服務的共同平台，將兩種電信網路的通訊服務、通訊網路、以及營運與商業模式結合為一個整體。在此平台上，固接式與行動電話的使用者能無阻礙地彼此通訊。

隨著通訊技術精進，電信自由化，以及通訊市場成熟，更多業者投入 FMC 的電信通訊市場。使用者有更多機會根據業者提供的服務項目、服務品質、以及收費機制選擇最適合的電信公司，更換電信服務公司的意願與機會也隨之提高。在此競爭激烈的電信環境中，能讓使用者即使更換電信服務公司，也依然能保有原來電話號碼的號碼可攜服務（number portability service）成為重要的服務項目。電信公司必須提供號碼可攜服務以吸引更多使用者，藉此提昇本身的市場競爭力。

號碼可攜服務實行之前，每個固接式或行動電話號碼可對應到使用者所屬電信服務公司的特定交換機，經由該交換機即可連結到該號碼對應的實際位址。然而，號碼可攜服務打散了電話號碼與電信公司以及使用者實際位址之間的關連性。為了將可攜式號碼重新對應至使用者的實際位址，電信公司必須建立一個共有的可攜式號碼資料庫（number portability database, NPDB）以管理 FMC 環境中

所有可攜式號碼與使用者實際所在位址間的對應關係，並委由一個中立的可攜式號碼管理機構（number portability administration center, NPAC）管理。電信公司經由查詢 NPAC 而將可攜式號碼轉換為使用者的實際位址，進而接通撥打給可攜式號碼的電話。

NPDB 的規模會隨著號碼可攜服務的使用者增加而擴大，當使用者數量很大時，NPDB 隨之增大，可能造成資料搜尋時間增加，使得整體可攜式號碼轉換的時間延長。然而，NPAC 的處理速度以及傳輸流量皆有限，為了進行可攜式號碼轉換所引發的大量 NPAC 查詢可能造成 NPAC 與網路壅塞，並阻礙其他話務進行。此外，自使用者撥出電話號碼開始，直到電話被接通，需佔用通訊頻寬進行接通話務的訊號傳輸。當撥打給可攜式號碼的電話數量增加，而可攜式號碼轉換的時間又延長，會降低通訊網路的使用效能與服務品質。

電信網路的架構設計使可攜式號碼的處理，以及號碼轉換時所需要的知識都集中在交換網路（switching network），要提高號碼可攜服務效能，必須降低交換網路壅塞的可能性；因此，必須減少 NPAC 查詢的數量才能提升號碼可攜服務的效能。根據這項觀察結果，我們提出將可攜式號碼轉換所需的知識分散到其他電信網路元件，由其他電信網路元件負責可攜式號碼轉換，藉此減輕 NPAC 的運算負擔以及所需處理的話務量。

根據這個想法，我們經由分析使用者的移動與通話行為，提出三項新的機制，以提升 FMC 環境中號碼可攜服務的效能：

1. 對於傳統固接式電話網路，我們提出將可攜式號碼轉換的知識與功能建立在 PBX 上。若大部分的可攜式號碼皆在 PBX 就被轉換為受話端的實際位址，NPAC 的查詢量將得到紓解，可攜式號碼查詢與轉換的時間縮短，可使號碼可攜服務效能提升。
2. 對於並未提供數據通訊頻道的 2G 行動通訊系統，我們提出的作法是以學校、辦公大樓、工廠等機構的區域網路為基礎，建立行動通訊的 PBX 系統，並將可攜式號碼轉換的知識與功能建立在這類行動通訊 PBX 上，以達到提

升號碼可攜服務效能的目的。

3. 對於具有數據通訊特性的 3G 行動通訊系統，我們提出的機制為：將個人化的可攜式號碼轉換知識存放於使用者端的智慧型手機，並利用 IP 網路更新手機上的知識，在使用者終端進行可攜式號碼轉換。此方法可將 NPAC 的查詢量減到最小、能最有效地紓解經由查詢 NPAC 以獲取受話端實際位址所造成的號碼轉換延遲。

在此論文中，我們證明經由上述的三項機制，可正確地將可攜式號碼轉換為受話端的實際位址，並可有效減少對 NPAC 進行可攜式號碼轉換的查詢量，降低可攜式號碼轉換所造成的延遲，提高號碼可攜服務的效能，並且可增進電信網路的使用效能與服務品質。所提出的方法對使用者以及電信服務提供者都有實際上的效益。

Enhancing The Efficiency of Number Portability Service On Fixed-Mobile Convergence Environment

Student: Ching-Wen Cheng   Advisor: Dr. Chyan-Goei Chung

Department of Computer Science
College of Computer Science
National Chiao Tung University

Abstract

Fixed and mobile convergence (FMC) is the combination of previously separate fixed and mobile services, networks, and commercial practices. A common platform to access both fixed and mobile telecommunications services is provided in FMC environment, such that users can set up calls to both fixed and mobile telecommunication systems.

Deregulation, market demand, and technological development encourage more service providers to join FMC telecommunications market. Users have more choices and are more likely to change service providers according to the service, quality, and the billing policy offered by the operators. Number portability (NP) service allows a user to keep the same telephone number when changing operators. In the competitive telecommunications market, operators must provide NP service to attract subscribers and to enhance their competitiveness.

However, NP service broke the relation between telephone numbers and the destination networks. For allowing switching systems to translate a ported number to the destination address, telecommunication operators establish a neutrally operated number portability administration center (NPAC) with a global NP database (NPDB) together to maintain the mapping of ported numbers and the information to reach the destination. NPDB maintains the mapping information of all ported numbers of the FMC environment. The size of NPDB grows enormously along the increase of NP users. The increase of NP users results in a huge NPDB will prolong the latency of data searching. The process and traffic capacity of NPAC is limited that the enormous NPAC queries for ported number translations may block other queries and cause congestion of the switching network. In addition, the communication resource is

occupied in call setup process. A large amount of NPAC queries will degrade the utility and service quality of the communication network.

Studying the architecture of telecommunications networks, the process and the knowledge of ported number translation is centralized in the switching networks. The efficiency of NP service can be enhanced by minimizing message passing for ported number translation. Hence, our approach is to dispatch the process and the knowledge of ported number translation to other network entities to alleviate the workload and the offered traffic of NPAC.

Based on the concept, we investigate the mobility and calling behavior of users and propose three new mechanisms for providing efficient NP service in FMC environment:

1. For fixed-line telecommunication systems, we propose to provide ported number translation functions in PBX-based telecommunications networks. When most NP calls can be translated in local telecommunications network, NPDB queries will be reduced and the efficiency of NP call setup will be enhanced.
2. For 2G mobile telecommunications system which does not provide data transmission channel, we propose an organization-based mobile telecommunications network to act as a mobile PBX to perform ported number translations in mobile local networks.
3. For 3G mobile telecommunications system which provides data communication features, we propose a mechanism to update routing information from NPAC to intelligent user terminals (e.g., 3G/WLAN dual-mode mobile handsets) via IP (Internet protocol) network. Thus, the routing information of ported numbers can be solved in user-end.

In this dissertation we prove that the amount of NP queries can be remarkably reduced by the above proposed mechanisms; hence, the efficiency and performance of NP service can be improved without affect the profit of telecommunications operators and users.

誌謝

能夠進入交大就讀，並且順利完成這篇論文，是幸運而快樂的經歷！最感謝的是我的指導教授鍾乾癸老師，在我碩、博士修業期間，不論學業或待人處世方面，始終不厭其煩地諄諄教誨。老師認真負責、人本思考的嚴謹態度,以及鼓勵與包容多元思考的氣度，始終為我表率。也要感謝師母吳秀琴老師一直不斷給予我關心與鼓勵。

更特別感謝饒仲華博士，在我就業與博士班就學期間，容忍我的莽撞與不成熟，引領我拓展研究與實務應用的新視野。

在此感謝撥冗參與口試審查會的林一平教授、陳耀宗教授、陳澤雄教授、楊竹星教授、以及蔡志宏教授（以上按姓氏筆劃排列）的指導與寶貴意見，皆是我往後需時時參考與警惕的原則。

感謝所有給予我指導與鼓勵的老師們，也謝謝甘苦與共的實驗室夥伴們，使我有機會成長，同時讓研究生涯增添許多樂趣。一直受到太多人的幫助，無法一一表示，在此一併致謝！

最後，衷心感謝家人對我的支持與鼓勵。謹將這份榮耀獻給我的家人，並致上最誠摯的謝意！

Index

List of Figures

List of Tables

Chapter 1. Introduction

Telecommunications technologies evolved from fixed-lined voice and data services to mobile multimedia communications. PSTN is a long developed and ubiquitous fixed-lined telecommunications system which has become a utility. The demand of mobility brought out the evolvement of mobile telecommunications systems. The popularity of mobile telecommunications carried out along the maturity of mobile technologies and the reduction of terminal and communication fees. The requirement of intercommunication between fixed and mobile telecommunications users causes the demand of fixed and mobile convergence (FMC). FMC supports fixed and mobile telecommunications services on a common platform. In FMC environment, subscribers of fixed-lined telecommunications systems must be able to set up calls to mobile telecommunications systems, and vice versa. Therefore, both the fixed-lined and mobile switching networks need to be able to determine and to route calls to the destination networks of the called parties.

Telecommunications liberation encourages more service providers to join fixed and mobile telecommunications market. Users have more choices and are more likely to change service providers according to the service, quality, and the billing policy offered by the operators. Conventionally, every operator has a unique numbering plan with respect to the national numbering plan and telecommunications policy. Changing telecommunications service provider implies changing telephone numbers. It is very inconvenient to users because they may miss calls that are set to the old subscription numbers. Number portability (NP) service allows a user to keep the same telephone number even when she changes operators. The calling and called parties will not sense about the physical routing of a NP call. In the competitive telecommunications market, operators must provide NP service to attract subscribers and to enhance their competitiveness.

The conventional telecommunications numbering plan connects a telephone number to a physical location or a definite subscriber. Switches of the core network route a call to the routing or destination address directly by parsing the prefix of the dialed number, the process is known as global title translation (GTT). However, NP service smashes the relation between dialed numbers and the destination routing addresses while ported numbers does not provide definite routing information for call routing. A mechanism to translate a ported number to the callee's current subscription number or the physical routing address is necessary of NP service. Hence, the facility for ported number translation must be available to all operator networks.

The facility for ported number translation can be embedded in switches (on-switch), or be maintained separately in the core network (off-switch) [1]. On-switch solutions modify routing rules in switches of core networks to route ported numbers to the destination address. The modification of routing knowledge in switches happens whenever a number was ported out or into a service operator, which is costly and inefficient. By contrast, off-switch solutions based on the architecture of intelligent network (IN) maintain the mapping of all ported numbers and the corresponding routing addresses of the numbers' subscription networks separately in a number portability database (NPDB). By accessing NPDB, switches can route ported numbers to the destination addresses of the called parties. Usually off-switch solutions are adopted because the update of routing information is easy without altering the switching system. Because dialed numbers and physical addresses are completely decoupled with number portability, a common information center to keep the link of a ported number and the current subscription network of the number is necessary. Usually all the telecommunication operators in the telecommunications environment establish a number portability administration center (NPAC) with a global NPDB together to fulfill the requirement. The NPAC must be neutrally operated. For the sake of load balancing and efficiency, operators may duplicate NPDB to local switching network to shorten the delay of NPAC queries.

In FMC environment, fixed-line telecommunications subscribers need to setup NP calls to mobile subscribers, and vice versa. The knowledge of translating fixed-lined and mobile ported numbers are maintained separately in fixed and mobile telecommunications networks. Therefore, the core network of every service provider in the FMC environment must be able to determine whether a dialed number indicates a fixed-lined or a mobile telecommunications network. The translation of ported numbers is processed in the caller's telecommunications network.

For providing NP services, two extra procedures are required in the call-origination process. First, whenever a number is dialed, the switching network needs to distinguish ported from non-ported numbers. In the IN-based NP service models, the determination of ported numbers requires the support of service switching and service data functions. Second, whenever a number is determined as a ported number, NPAC queries are triggered to translate the numbers to a reachable address. The two procedures will elongate the call setup time and increase the traffic load of the switching network. Moreover, not only ported numbers trigger NPAC queries. When a number was recorded as a ported number, the entire group of numbers was taken as ported numbers [2]. Every such NP and non-NP call brings about ported number translation and queries NPAC. As shown in Fig. 1-1, the design of telecommunication networks assembles the routing and signaling processing knowledge in the central switching center. For providing NP service in FMC environment, NPDB must maintain the routing information of all ported numbers of fixed-lined and mobile telecommunications systems. Whilst the size of NPDB grows enormously, searching routing information in a NPDB will be time-consuming. When the offered load of NPAC queries exceeds the threshold workload of NPAC (the dot line in Fig. 1-2), the waiting time of NPAC queries increases rapidly because of the queuing and database searching delay (Fig. 1-2). The latency delay of routing information query may block other queries and cause the congestion of the transmission network, which results in prolonged response time of ported number translation.

Some researches tended to enhance the efficiency of NP services by shortening the search delay of NPDB queries. [11][11] proposed to enhance the efficiency of NPDB queries by improving searching algorithm. However, bandwidth is scarce resource of telecommunications networks, which will be occupied during the process of call setup. The prolonged call setup delay due to ported number translation will lead to extra operation time and bandwidth consumption; moreover, subsequent calls will be blocked for the scarcity of bandwidth. Operators must bear the expenditure on extra consumed communication resources without bringing operators any revenue. Enhancing the efficiency of NPDB queries is not sufficient to mitigate the heavy traffic load.

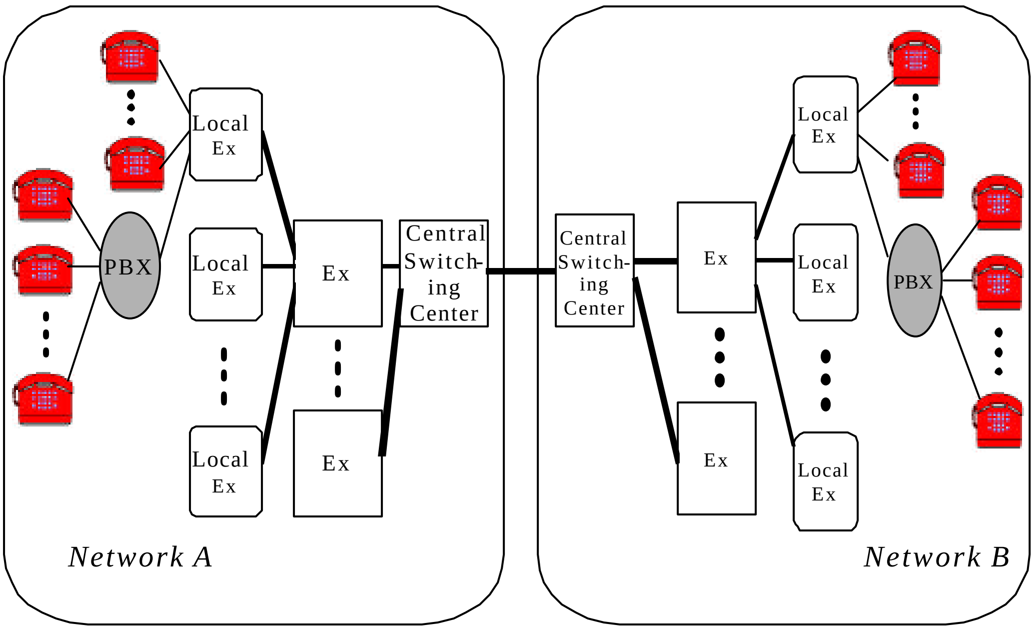

Fig. 1-1 The approaches of applying caches to telecommunication networks

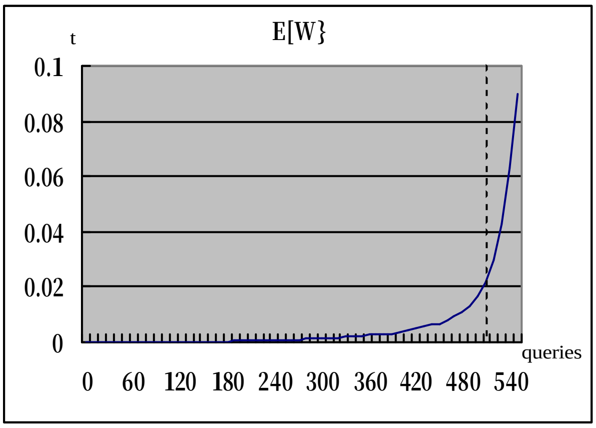

Fig. 1-2 The bottleneck of NP service

The study of [5] discovered that 99% of the calls are set to the numbers had been called in a week. That is, most of the accessed data will be accessed again in the near feature. Based on the observation, keeping the recently accessed data bcally can avoid a large part of long-term NPAC queries and shorten the average NP call setup delay time. Accordingly, [4][5]suggested applying caches to operator networks can effectively alleviate the amount of query messages and improve the efficiency of information query. The effect of caches depends on the cache hit rate and the cache size [5]. The amount of served users of central switches is so large that the required cache size should be very large too. An enormous cache may incur longer search time; hence, the numerous requests of ported number translation will congest the core network of telecommunications networks. On the contrary, the size of caches in local switches is small but very helpful to enhance the efficiency of ported number translation [3]. But the exchange of updated routing information between NPDB and local caches must be transmitted through voice transmission lines, which will consume extra communication bandwidth and crowd out arrival calls.

The works of [6][7] presented the chance to solve NP problem by caching routing information of ported numbers in user terminals. The results showed that shift knowledge to intelligent peripherals with better computation power and storage capacity can minimize information passing and effectively alleviate traffic load of core network. But replacing user terminals by intelligent peripherals is a tremendous project. On the other hand, the correctness of the routing information of dialed ported numbers must be guaranteed to set every call to the right destination. Consequently, a mechanism to synchronize the routing information of NPDB and user terminals is necessary. However, PSTN and mobile telecommunication systems before 2.5G (GPRS) were not designed for data transmitting. Distributing information from core network to user-ends is not easy in PSTN and the mobile telecom systems before 2.5G. The telecommunications systems must seize lines or arrange bandwidths to dispatch information from core networks to customers. Although data channels are available in 2.5G and 3G systems, delivering routing information to user-ends will consume a large amount of computation resources and bandwidth. Calls will be blocked during the dispatching of information, the cost is too expensive to be feasible.

From the above discussion, we found that the existing solutions confront the following problems: (1) non-intelligent peripherals can not translate ported numbers to the destination address of the called party. NP calls originated from non-intelligent peripherals require NPAC queries for ported number translation. NPAC will be encumbered with considerable quantities NPAC queries that the performance may be degraded and block other queries. Thus, the call setup delay will be prolonged. (2) The update of routing information in intelligent peripherals is transmitted through telecommunication lines. That may occupy telecommunication resource and obstruct call setup.

Considering the first problem, a mechanism to determine and translate ported numbers dialed from non-intelligent peripherals in the early stage of NP call process is the key factor to alleviate the workload of NPAC and to prevent the congestion of telecommunications bandwidth. More NP calls be solved and translated in the early stage, less routing information queries are issued. Thus, NPAC query delay will be shortened and the network congestion will be prevented. On the other hand, the update of the routing information of ported numbers on user terminals can be performed through data networks (e.g., Internet Protocol-based network) to omit unnecessary telecommunications resource consumption. These two notions can support service providers to decrease the cost of providing NP service and help subscribers to reduce the waiting time of NP call setup. Based on the concept, three new mechanisms for providing efficient NP service in FMC environment are proposed in this research:

1. For fixed-line telecommunication systems, we propose to provide ported number translation functions in PBX-based telecommunications networks. When most NP calls can be translated in local telecommunications network, NPDB queries will be reduced and the efficiency of NP call setup will be enhanced.
2. For mobile telecommunications systems before 2.5G that do not provide data transmission channel, we propose an organization-based mobile telecommunications network to act as a mobile PBX to perform ported number translations in mobile local networks.
3. For 2.5G and beyond 2.5G mobile telecommunications systems that provide data communication features, we propose a mechanism to update routing information from NPDB to intelligent user terminals (e.g., 3G/WLAN dual-mode mobile handsets) via IP network. Thus, the routing information of ported numbers can be

solved in user-end.

The amount of NP queries can be remarkably reduced by the above proposed mechanisms; hence, the efficiency and performance of NP service can be improved without affect the profit of telecommunications operators and users.

A brief synopsis of the remaining chapters follows.

- Factors which affect the efficiency of NP call setup process: In chapter 2 we state the state of the art of NP solutions. The researches about solving the problems are reviewed, and the factors which affect the efficiency of NP service in FMC environment are investigated.
- Enhancing the efficiency of PSTN NP service: PSTN is the most important worldwide fixed-lined telecommunication system. By observing the calling behavior of PSTN users, it is found that most of the calls are originated from PBX of organizations in business hours. In chapter 3 we introduce a mechanism to perform ported number translations in PBX to shift the workload of NPAC, and discuss the feasibility and cost-efficiency of the mechanism.
- Enhancing the efficiency of mobile NP service before 2.5G mobile telecommunication systems: In mobile telecommunication systems without the support of data transmission channels, solving ported number translation in user terminal is not feasible. In business hours, most of the mobile users reside in designated areas (usually in the organization they belong to) and most of the calls are originated from organization members. In chapter 4 we introduce a mobile PBX to achieve the purpose of providing ported number translation in organization-based networks to enhance the efficiency of mobile NP service. The architecture, functions, and the operation model of the mobile PBX are illustrated. Also the benefit of the mechanism is studied.
- Enhancing the efficiency of mobile NP service in 3G and beyond mobile telecommunication systems: In the mobile telecommunication environment with the support of data transmission channels and customized data service, providing the function of ported number translation in user terminals is practicable. In chapter 5 we present a mechanism for user terminals to convey the translated routing information of ported numbers, and for the switching network to distinguish the information and route the call. An IP-based data synchronization scheme is offered to guarantee the validity of routing information, and an algorithm to decrease the power consumption of user terminals is proposed.
- Finally, conclusions and future works are drawn in chapter 6.

Chapter 2. Background and related works

2.1 Background

Number portability (NP) is a generic service concept which provides a network capability to enable a subscriber to keep his/her telephone number the same with a change of network operator, location and service type, etc. There are three types of number portability services: service provider portability, location portability, and service portability [13]. With service provider portability, a subscriber may switch service provider without changing his/her telephone number. With location portability, a subscriber may change location without changing telephone number. With service portability, a subscriber may keep the same telephone number when changing telecommunications services, such as changing from fixed-lined telecommunications service to mobile service.

In most countries, location portability and service portability are not enforced, and only service provider portability is implemented. Service provider portability is considered essential for fair competition among operators, while location portability and service portability are typically treated as value-added services. Location mobility has been implemented in mobile system because whenever a subscriber moves into a mobile network, the *visiting network* updates the location information of the subscriber to the subscriber's subscription network. Besides, the numbering plans of fixed-lined and mobile telecommunication services are different in most countries, where service portability is not available unless the numbering plan is modified. Therefore, we focus our discussion on service provider portability.

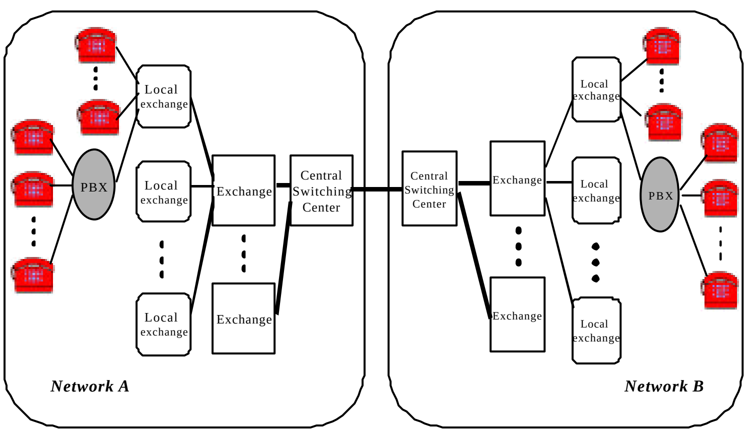

Fig. 2-1 The hierarchy of telecommunication networks

As illustrated in Fig. 2-1, a telecommunications network is a hierarchical architecture consisting of several layers of exchanges. Usually the North American Numbering Plan (NANP) in the format of NPA-NXX-XXXX is adopted as the naming mechanism of a fix-lined telecommunications system, where numbering plan area (NPA) is a non-geographic code or a service access code, N is a number between 2 and 9, and X is a number between 0 and 9. Following NANP, a local exchange serves the numbers from NPA-NXX-0000 to –9999. The communication region of a city may consist of several local exchanges. Once a number within a local exchange was ported to another network, the local exchange considers the whole set of numbers, which is called a number block, as ported numbers.

Conventionally, every operator has a unique numbering plan with respect to the national numbering plan and telecommunications policy. In fixed-lined telecommunications system, every telephone number indicates a physical location. The routing of calls relies entirely on the network that originally issued the phone number, which is called the *donor network* of the telephone number. For routing, number portability relies on the capability of a switching network to route a ported number to the network that is currently serving the number. In mobile telecommunications system, a *directory number* (mobile telephone number) indicates a subscriber, while the *identification number* uniquely identifies a mobile station (MS) in the mobile network. The *donor network* which first issued the directory number to a subscriber, and the *subscription network* which a subscriber registered to must track the location of the subscriber and his/her MS. In order to distinguish ported numbers and determine the destination network of them, number portability is a necessary network function to allow the switching network to route calls to the called parties.

NP implementation schemes can be classified into on-switch solutions and off-switch solutions [14]. In the fixed-lined network systems, for intercepting and routing ported calls efficiently, the on-switch solution always routes calls to the donor network and are then onward routes them to the destination network. Analogous to the on-switch solutions of the fixed-lined telecommunication system, signal relay function (SRF)-based solutions modify switches to support NP service in mobile network systems. SRF-based solutions founded on SS7 communication architecture enhance the switch functions and utilize the MAP (mobile application part) protocol to enable the translation of the dialed number and the destination address. Depending on the implementation, the translation can be performed in the donor network or the subscription network. The SRF is typically implemented on signaling transfer points (STP) in the SS7 communication model.

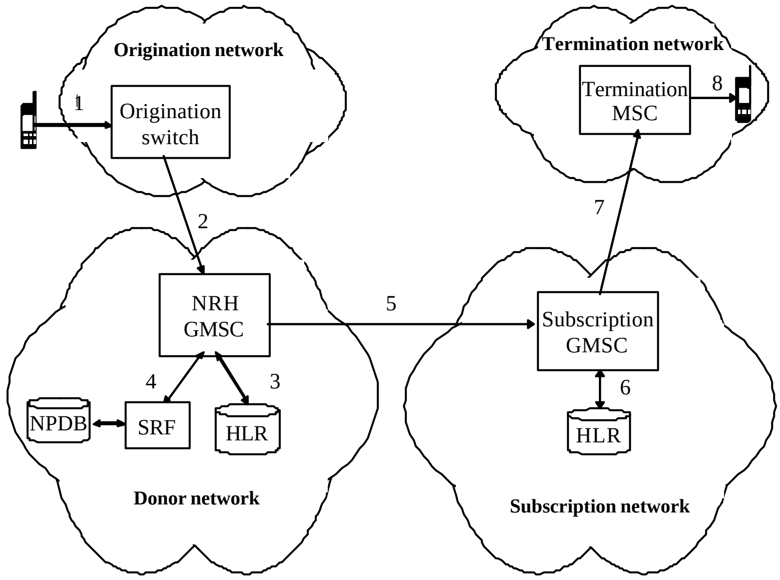

Fig. 2-2: SRF-based NP call routing

Methods of on-switch solutions implement routing knowledge on the switching center of service providers. When a caller dialed a call via an *originating network*, which a caller connects to, the originating network routes the call to the *donor network* by the prefix of the dialed number. The call is routed via the donor network to the destination network, hence the mapping of dialed numbers and the routing addresses of the gateway of the destination network or the destination address of the called party are maintained in the gateway switches of the donor network. The simplified NP call process is illustrated in Fig. 2-2. The origination network receives a call initiation request from a subscriber (step 1). It identifies the donor network of the called party by the prefix of the dialed number (MSISDN). The origination network issues an ISUP IAM message to the donor GMSC to initiate a call (step 2). The donor GMSC consults HLR and identifies the number was ported out (step 3), then it consults NPDB by MAP sending routing information message (step 4) to determine the routing number of the subscription network. The donor network forwards the IAM message to the subscription network (step 5), and the subscription GMSC queries HLR for the routing number (mobile station roaming number, MSRN) of the called party (step 6). The MSRN indicates the address of the termination switch, thus the subscription network can route the IAM message by the MSRN to the termination switch to set up the call (step 7).

Following this method, the operation logic of a switching center alters whenever a number is ported out or in. On-switch solutions confront the following problems: (1) The frequent alteration decreases the stability of communication services and increases the cost of system operation and maintenance, and (2) the growth of ported numbers leads to a large number of routing messages for number translation.

On the contrary, to prevent the alteration of switch networks, off-switch solutions use Intelligent Network (IN) to intercept and route ported calls without the participation of the donor network. [15]. NPDB that manage the mapping of ported numbers and corresponding routing information are involved in the process of call setup. Switch network queries NPDB to obtain the routing address of the dialed numbers that are marked in the switch as ported. NPDB can be established as an internal database that keeps only the NP information of numbers which are assigned from or subscribed to the network, or be centrally maintained by a neutral organization (Number Portability Administration Center, NPAC) that all ported numbers of every service provider are recorded. Fig. 2-3 represents the architecture of IN-based NP services in a switched circuit network (SCN, including ISDN and cellular networks) and IP telephony interoperable environment. In IN-based network architecture, the service data functions (SDF) return a routing number for a ported number that indicates an end-point or an end-user in a network. The service management system (SMS) manages the content of SDF and handles the data consistency between SDF and NPDB. For the query to the IN nodes, different protocols are used in different networks and by different operators. In the IP network, number gateways translate telephone numbers to IP addresses. The location servers (LS) work as SDF that maintain the location information of subscribers.

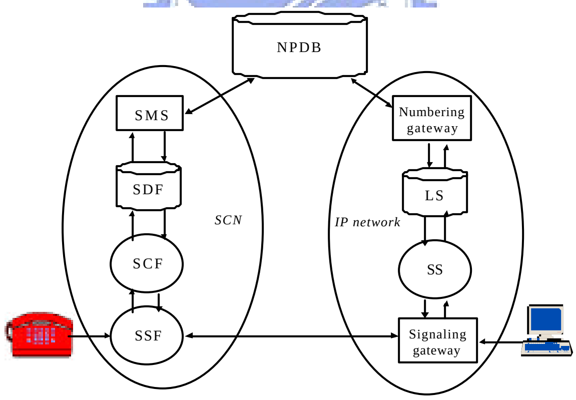

Fig. 2-3 IN-based number portability service architecture

As shown in Fig. 2-4, the origination network receives a call initiation request from a subscriber (step 1), it determines the dialed number indicates to a NP subscriber and issues an ISUP IAM to the donor network for initiating a call (step 2). The donor network queries HLR and determines the number was ported out (step 3), and consults NPDB by an INAP initialDP for the routing number of the subscription network (step 4). The donor network forwards the IAM message to the subscription network (step 5). The subscription network queries HLR for the MSRN of the called party (step 6) to reach the termination network, and forwards the IAM message to the termination MSC to setup the call (step 7).

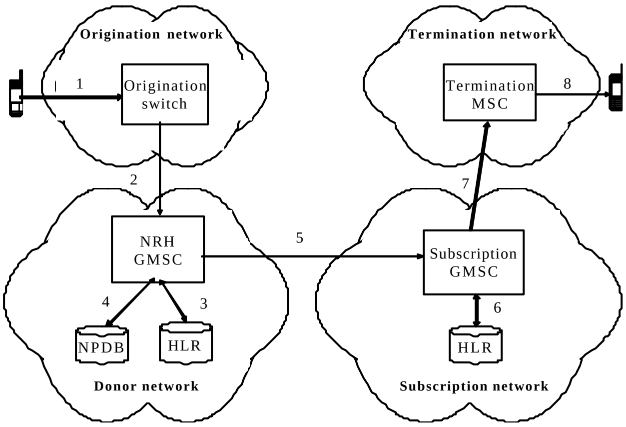

Fig. 2-4 IN-based solution of NP call routing

The IN-based solutions differ form the SRF-based solutions in the way to access NPDB. In the SRF-based solution only GMSC can query NPDB. But the IN-based solutions are implemented in the service control point (SCP). Every switch equipped with the IN protocol can access NPDB [15]. SRF-based solutions may centralized workload to and burden specific network entities, where IN-based solution mitigates that problem. By IN architecture, it is not necessary to re-test the function of switches when updating the routing information of new ported numbers. Off-switch solutions are wildly adopted for better flexibility and extensibility.

There are four off-switch schemes for supporting NP service: all call query (ACQ), query on release (QOR), call dropback (also known as return to pivot, RTP), and onward routing (OR) [16].

- All call query (ACQ): ACQ scheme requires a centralized database to keep information of all ported numbers of every service providers. As shown in Fig. 2-5(a), the originating network detects the dialed number is a ported number, and initiates a query to NPAC. NPAC returns the routing information of the dialed number to the originating network, then the originating network routes the call to the destination network to set up the call. The determination of ported numbers is performed in the origination network. In this method, the operation of the donor network was not affected by the operation of NP call setup. ACQ is the most efficient of using the network transmission facilities; therefore, it is widely adopted as NP solutions in many countries.
- Query on release (QoR): QoR scheme is a form of call re-routing method, which grounds on the release messages of the donor network that detect a dialed number was ported out. In Fig. 2-5(b), when the originating network detects that the dialed number is a ported number, it routes the call to the donor network by the prefix of the number. If the number is ported out, the donor network returns a release message to the originating network, then originating network queries NPAC for the routing information of the dialed number. When the originating

network receives the routing information, it re-routes the call to the destination network. The occupancy of transmission resources during the routing of calls may degrade the efficiency of source network.

- Call dropback: In Fig. 2-5(c), the originating network receives a call from the caller and routes the call to the donor network by the prefix of the number. The donor network determines the number was ported out, it returns the routing information of the number to the originating network and release the call. Then the origination network re-routes the call to the destination network. Following this method, the routing information is maintained separately in different donor networks of ported numbers, and the donor networks of every ported number need internal NPDB for recognizing whether the number is ported out.
- Onward routing (OR): As pictured in Fig. 2-5(d), the originating network routes the received call to the donor network of the dialed number. The donor network detects the number was ported out, and checks an internal NPDB for the routing information. Then the donor network routes the call to the destination network by the routing information. The scheme requires the setup of two physical call segments, one from the originating network to the donor network and the other from the donor network to the destination network. It is the least efficient in terms of using the network transmission facilities.

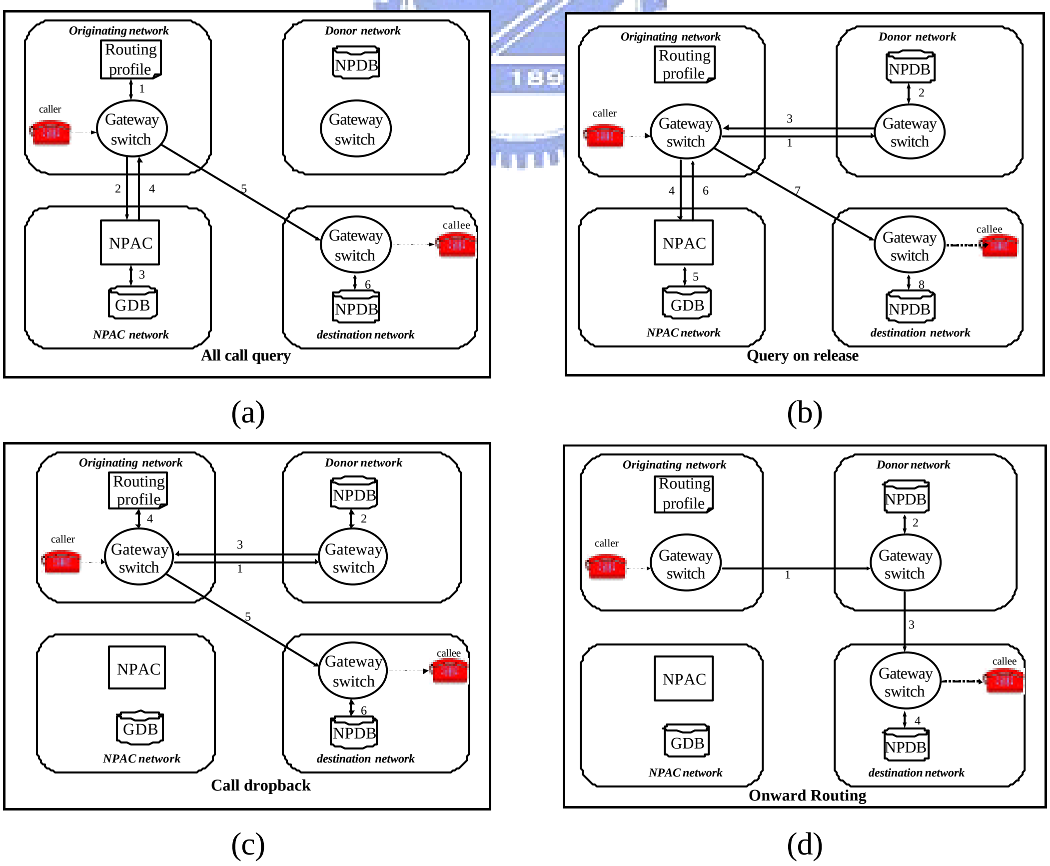

Fig. 2-5 Off-switch NP schemes

The four schemes exist in the NP solutions of different countries: UK, Finland, France, Germany, Span, Singapore, etc. [17][18][19] The considerations of which scheme to adopt include the network resources, the policy of addressing and routing, the impacts on the signaling system, and the interworking with other services. ACQ and call dropback are the two most popular solutions.

Providing NP service in fixed-lined telecommunication environment requires two procedures to determine and translate ported numbers in call setup process. First, whenever a number is dialed, the switching network needs to distinguish ported from non-ported numbers. Every call in FMC telecommunications environment can be set to a ported or non-ported number of a fixed-lined or mobile telecommunications subscriber. The switching network of a telecommunication system must determine the destination network of a dialed number to setup a call. While NP service broke the relation between dialed numbers and the corresponding subscription networks, the ported number marks of both fixed-lined and mobile telecommunication networks need to be maintained in every network in FMC environment to distinguish NP and non-NP call processes. The ported number marks in switches increase with the growing amount of NP users. When the amount of NP users grows to be enormous, the latency of this procedure will be long. Second, every call terminated to a ported number initiates NPDB queries. Moreover, not only ported numbers trigger NPDB queries. When a number was recorded as a ported number, the entire group of numbers was taken as ported numbers [15]. Every such NP and non-NP call brings about queries to NPDB. The databases search time is log_2N in average for every call, where N is the number of NP users. As shown in Fig. 2-1, the hierarchy of telecommunications networks assembles the routing and signaling processing knowledge in the central switching centers. When the arrival rate of NPDB queries exceeds the threshold service rate of NPDB (the dot line in Fig. 1-2), the waiting time of NPDB queries increases rapidly because of the queuing delay (Fig. 1-2). The latency delay of NPDB searching will block other NPDB queries and cause the congestion of the transmission network, which results in prolonged response time of ported number translation. The prolonged call setup delay due to ported number translation leads to extra operation time and bandwidth consumption. Operators must bear the expenditure on extra consumed communication resources without bringing operators any revenue.

The process to set up a NP call consists of the following procedures: process and transmission of messages, determination of NP calls, NPDB queries for the translation of ported numbers, and seizure of transmission line. Let t_{local} and t_{global} represent the transmission time in local and in global network, $t_{process}$ and $t_{seizure}$ are the time required for the switching network to process messages and allocate transmission line/channel to the call, $t_{determination}$ is the latency for distinguishing portable and non-ported numbers, and $t_{NPDB\text{-}query}$ presents the routing information query delay when consulting NPDB. The setup delay of a NP call can be presented as the following:

$$t_{call\text{-}setup} = t_{local} + t_{global} + t_{process} + t_{seizure} + t_{determiniation} + t_{NPDB\text{-}query}$$

Where $t_{translation}$ and $t_{NPDB\text{-}query}$ will increase when the amount of NP subscriber increases; however, t_{local}, t_{global}, $t_{process}$, and $t_{seisure}$ will not vary with the amount of NP subscribers. Therefore, data search delay must be reduced to enhance the efficiency of NP service and to utilize telecommunication resources.

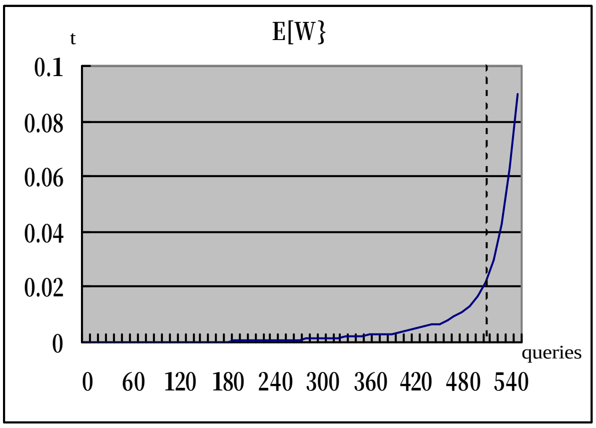

Fig. 2-6 The bottleneck of NP service

2.2 Related works

For the purpose of enhancing the efficiency of NP service, some researches tended to enhance the efficiency of NP services by improving the response time of NPDB queries. [11][11] proposed to enhance the efficiency of NPDB queries by improving searching algorithm. However, bandwidth is scarce resource of telecommunications networks, which is occupied during the process of call setup. When data search delay exceeds the threshold, the efficiency of NP call process degrade rapidly that will result in the congestion of transmission lines. Enhancing the efficiency of NPDB queries is not sufficient. Unless service providers mitigate the heavy traffic load or establish more telecommunications lines to tolerate the traffic load, the enhancement is limited.

In order to alleviate the heavy workload of operator networks to enhance the efficiency of NP service, many studies proposed that implementing caches in telecommunications system can effectively alleviate the amount of query messages and improve the efficiency of data access [16][4][5][20]. Kim and Yong stated the factors affecting the cache hit ratio in mobile computing environment including the distribution of queried target objects and the query pattern [16]. Telecommunication networks are designed as computation and intelligent centric, implementing caches in operator networks is easy to maintain and benefit environment-dependent decision making [22]. Refer to the hierarchy of telecommunications network (Fig. 2-1), central switches serve so many users and the dialed numbers a central switch receives is scattered. Every NP call requires ported number translation. The size of a cache should be large to accommodate sufficient data (double of the size of the numbers dialed from users) to achieve acceptable hit ratio. Jain et al. proposed a hashing scheme in [11] to improve cache hit ratio. But keeping the large amount of portable users in FMC environment will require a lot of memory size. The cache size and the probability of collision will increase as the amount of portable users increases. Chan and Leong addressed that clients should take a more active role in maintaining cached items. In [21], [16] and [4], authors suggested distributing spatial replicas of databases to different sites and proposed caching schemes on mobile handsets (MS) with respect to environment properties to provide efficient data access to users. However, mobile users move in and out of several service regions, the limited cache size on a MS will be hard to keep ample environment information. In order to guarantee the validity of cached data, MS need to update cached data frequently according to temporal and spatial properties. The requests of cache updates from a large amount of MS will cumulate the traffic of the operator network.

The study of Carpenter et al discovered that 99% of the calls are set to the numbers had been called in a week, and assumed that an individual customer's calling behavior exhibits a strong locality of reference [5]. Assuming intelligent peripherals are available, they proposed to maintain a profile and a cache of a user's frequently accessed data in the user's terminal. The work of [6][7] presented the chance to solve NP problem by enhanced user terminals. The results showed that shift knowledge to intelligent peripherals with better computation power and storage capacity can minimize information passing and effectively alleviate traffic load of core network. Therefore, performing ported number translation in user terminals will alleviate the traffic load of NPDB queries and mitigate the workload of NP call process in operator networks. But it is expensive and tardy in updating user terminals comprehensively. And the update of the routing information in local caches will occupy telecommunications lines, which consumes extra communication bandwidth and crowd out arrival calls.

While a comprehensive solution to solve NP problem is not available, another solution is to find some mechanism which can effectively reduce the amount of requests needed for ported number translations. If the amount of requests which need for ported number translation is small, the switch system will have enough computation capability to handle the received call request. From the result of [5], number translation should be performed in a network entity with the properties of evident dialed number locality; consequently, the quantity of routing information needed to be maintained is modest. In addition, the network entity must possess ability of storage and computation to keep valid routing information of ported numbers and to provide the service of ported number translation.

In the fixed-lined telecommunications environment, most organizations establish PBX to save telecommunications cost and benefit intra-organization communications. That is, most of the calls (both NP and non-NP calls) in business hours are relayed by PBX to the public telecommunications network. PBX is a network entity with computation power and storage that can perform the function of ported number translations. In business hours, a major part of the calls are generated from organizations. Applying the knowledge of number translation and keeping the routing information of ported numbers in organization-based telecommunication networks will be an effective approach to enhance the efficiency of NP service. In mobile telecommunication systems before 2.5G, intelligent terminals and data transmission channels are not available. Ported number translations can not be performed in user terminals. An organization-based network which possesses the property of dialed number locality can be utilized to provide the service of ported number translation. However, mobile PBXs are not generally available. A network entity to act as a mobile PBX in the mobile telecommunication network is required. In 3G mobile telecommunication systems, user terminals are smart and powerful to perform computation and storage tasks. Because the dialed numbers of an individual user often presents strong locality, keeping the routing information of a user's frequently dialed ported numbers in the terminal will benefit the efficiency of NP service.

On the other hand, a mechanism to synchronize the distributed routing information with that in NPDB is necessary to guarantee that every dialed ported number can be translated to the right destination address. Dispatching routing information from NPDB to organization-based network and 3G mobile terminals can be transmitted by IP-based networks without consuming telecommunication transmission resource. When ported numbers are resolved in the early stage of NP call process, the translated routing information must be recognizable to the public switching network. A method for local telecommunications networks and 3G user terminals to notify the switching core networks that the dialed numbers were translated to effective addresses is also needed.

Based on the above idea, the following three mechanisms which can tell the central switching system that the call-origination request needs for ported number translation or not are proposed:

- PBX with ported number translation capability

- An organization-based mobile PBX system with ported number translation capability
- The dual mode mobile phone with ported number translation capability

The details of these three mechanisms are described in the following three chapters.

Chapter 3. An organization-based cache approach for supporting fixed-lined telecommunications number portability service

3.1 Motivation

In the conventional fixed-lined telecommunication system, to connect a call, the PSTN (public switched telephone network) switching network should translate the global title digits (i.e., the dialed number) to determine the destination network and the routing address (GTT), and a routing table is used in translation. Number portability service break the relation of a PSTN telephone number and destination network. The switching network has to determine whether the dialed number is ported or not and translate the ported number to the corresponding global title digits before using the GTT process to connect the call. Hence the waiting time of the call setup should be increased, and the call setup waiting time will increase when the size of NPDB is increased. Since every dialed rumble should be checked, the amount of call connection per unit time of the central switching network will be decreased. Of all calls are dialed from intelligent terminals, the central switching network doesn't need to do ported number checking and translation because the received global title digits is correct. However, it is impossible in real situation. We should find some ways to reduce the number of calls which the central switching network should do ported number checking and translation activity.

A major part of conventional telecommunication traffic load was generated from organizations (i.e. government offices, enterprises, factories, etc) during the office hours. Most of the calls (both NP and non-NP calls) in business hours are issued from organizations. Organizations usually establish PBX (Private Branch Exchanges) for saving telecommunication cost and benefiting intra-organization communications. That is, most of the calls (both NP and non-NP calls) in business hours are relayed by PBX to the public network.

The telephone numbers dialed from an organization often exhibit locality. For example, organizations like government departments or retail businesses have static contact targets to solve problems or to supply merchandise; manufacturing industries have a set of upstream suppliers and downstream customers, and insurance companies have a set of static cooperation enterprises, such as banks, airlines, hospitals. We found that almost every organization has a set of frequent dialed numbers that consists of businesses partner, cooperation industries, friends and families of employees, etc, and the variation of these contact targets is infrequent. Of these frequent dialed numbers and corresponding routing information is kept in the memory of PBX, then when one of these number is dialed PBX can tell the switching network that this number need not do number checking and translation activity. This will save lots of connection loading of switching network. Based on this concept, the design of PBX with frequent dialed number routing information is described as follow.

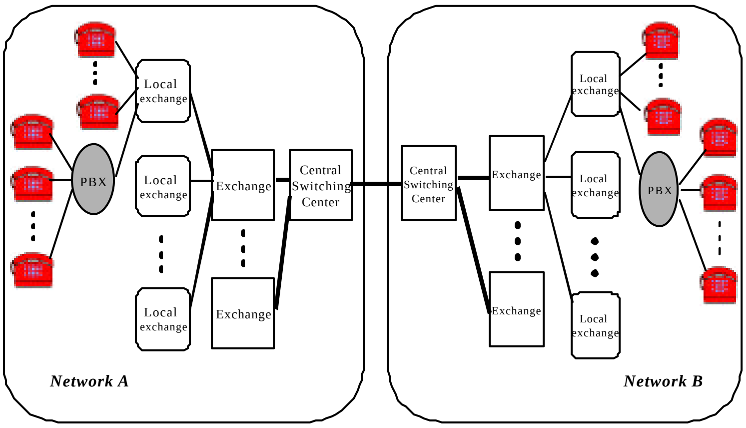

Fig. 3-1 The hierarchy of telecommunication networks

3.2 Applying caches to PBX-based networks

A PBX is a telephone exchange which provides internal communications for a set of users on local lines while allowing all users to share a certain number of external phone lines. Every PBX belongs to an organization that enables the communication between organization members without going through the public network, and to make and receive calls from subscribers served by the public network. A PBX connects to the public network by one or more physical channels. The communication between a PBX user and a subscriber of the public network is routed by the public switching network which the PBX connects to. While a PBX may not have the capability to perform database search or processing functions, a computer is added to the system by the open application interface (OAI) as shown in Fig. 3-2 [5]. Hence the computation tasks are realized on the PBX.

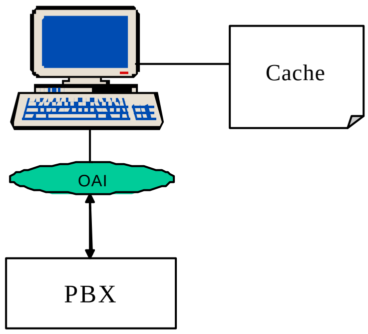

Fig. 3-2 An OAI-enabled PBX

The communication beyond a PBX service region is transmitted through the public network. In the conventional call originating process, a PBX routes every external call to the connected public switching network, and the switching network follows the embedded routing logic to route the request. When applying cache-stored ported number translation knowledge to PBX, a PBX checks cache before routing an external call to the connected local exchange. A cache hit indicates that the routing address of the dialed number is determined and confirmed, which is proved by the relaying PBX, and the NPDB queries can be omitted. Consequently, a mechanism is required for a PBX to inform the public switching network about the confirmation of the routing address. Thus, the switching network can omit NPDB queries and route the call to the address directly.

According to the function of special code service in IN-based network, a PBX can add a special code (e.g., *14*, *30*) in front of the dialed number to indicate that the routing address of a dialed number is appended to the call originating request. The switching network recognizes the code and routes the call by the appended routing address directly.

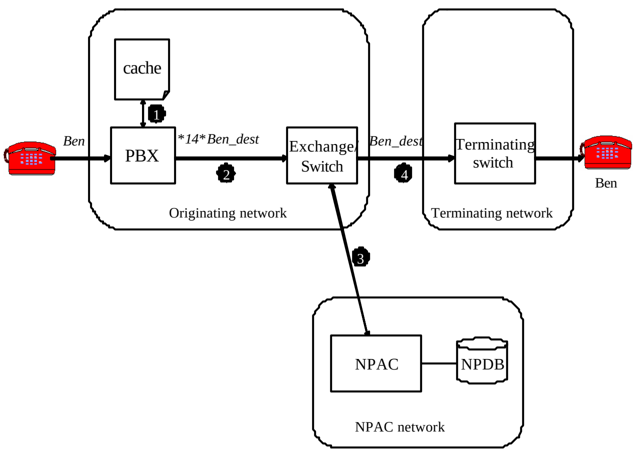

Fig. 3-3 A special code (e.g., *14*) is used to indicate a cache hit

Fig. 3-3 illustrates the basic NP call origination process. A call is set from a member within a PBX service region to a public telecommunications subscriber *Ben*. The PBX determines the call is an external user and checks cache for the routing address of *Ben* (step 1). In the case of cache hit, the PBX obtains and appends the routing address of Ben to the call originating request, adds a special code "*14*" in front of the request, and sends the request to the public network (step 2). According to the routing address, switches routes the request to the terminating network (step 4) without querying NPDB. The terminating network routes the call to *Ben* to setup the call.

When the dialed number and the corresponding routing address are not kept in the cache, it will be a case of cache miss. Hence, the PBX routes the call originating request to the public switching network without modification and special code. The public switching network determines the dialed number is a ported number consults NPDB to translate the ported number (step 3), then routes the request to the terminating network according to the obtained routing information (step 4).

The signaling systems of service providers usually have the function to process special codes (e.g., the code "*67" in USA prohibits displaying the caller's telephone number to the called party). Applying caches to PBX requires adding a routing rule in the signaling system without altering existing operation logic. It is feasible for service providers to diagnose and process the special code to omit time-consuming NPDB queries.

3.2.1 Issues

From the perspective of organizations, cost is the most important issue when applying caches in PBX. The cache size should be small but sufficient for comprehending frequently used data that can perform fine cache hit rate. The update of routing information in a cache should be simple and efficient without obstruct the communication service of an organization. From another point of view, routing information of ported numbers is valuable information of telecommunications service providers that should not be distributed to subscribers arbitrarily. The distribution and update of routing information should be based on contract or agreement without interfering with the communication service. Thus, the issues of applying caches to PBX encompass the policy of establishing a cache, the update of cached data, and the size of a cache.

- **Cache establishment**

Caches on PBX should keep as many routing information of ported numbers as possible to minimize NPDB queries. However, the size of a cache is restricted that the data can be kept is limited. Cache hit rate increases when the accesses of data exhibit locality. In order to improve the hit ratio of cached data, the cached data should expose the communication habits of users in the PBX service region.

There are two approaches for cache establishment: Cache the most recently dialed numbers because they might be dialed again in a span; or cache the frequently dialed numbers for they are the most usually dialed numbers for a long-term observation. The two approaches are referenced as dynamic cache policy and static cache policy respectively in the following descriptions.

Under the assumption that the recently dialed numbers are most likely to be dialed repeatedly in a span, the dynamic cache policy argues to keep the most recently dialed numbers. When the cache is full and a new data entry is being inserted, a replacement process is triggered automatically by the cache management system to replace an old entry with new one. Dynamic cache policy has the advantage that the cached data represents the recent calling behavior of users. However, considering that a cached number is ported to another service provider, the cached routing information becomes obsolete. A hit of obsolete data brings about a miss-routing that consumes extra signaling and transmission resources and prolongs the call setup time. Calls that arrived before the obsolete data was updated will be routed to wrong routing addresses or be blocked until the data was renewed. To update obsolete data in the cache dynamically, the public switching system must be modified to notify PBX that the cached data is wrong and to send back the renewed routing information immediately. To prevent miss-routing caused by obsolete data, cache update must be processed immediately when a data-renewal notification is received; furthermore, cache replacement must be processed whenever a new data is inserted to the cache. Access an updating cache will result in miss-routing. For the reason, cache access should be forbidden during cache update. But all the external calls will be blocked in the period. The costs to modify the signaling system of a public telecommunication network, to management a dynamic cache, and to update renewed routing information to the cache dynamically are expensive. Consequently, dynamic cache policy is not a feasible solution for PBX-based caches.

On the contrary, a static cache keeps the telephone numbers and the corresponding routing addresses of the most frequently contact targets of a PBX service region, the telephone numbers are referenced as frequently dialed numbers (FDN) that represents the cooperative enterprises, agents, families and friends of organization members, etc. The establishment and alteration of a static cache are manually performed by the system administrator. When a member joined an organization, the member proposes a set of FDN to the system administrator. The system administrator sends the FDN set to a contracted telecommunications service provider to obtain the routing information of the FDN set. Service providers provide an add-on service to allow PBX to consult the routing information of telephone numbers, and to notify PBX of the update of routing information. When an organization telecommunication network for the service, the telecommunications network maintains a profile of the PBX's FDN of the organization. Hence, the telecommunication network can notify the PBX to update cached data when a number in the profile changes subscription network, then the PBX will query the service provider to often the renewed routing information.

The query of routing information is a batch process that can prevent inconsistent data in the cache. In order to prevent the block of calls during the update of cache, routing information exchanges between PBX and the subscription network should be transmitted by IP-network in off-time to reduce the impact on the telecommunications resource. While the dialed numbers of an organization have locality, and the change of members is infrequent, the variation of FDN is gentle. The complexity of establishing and maintaining a static cache is much less than that of a dynamic cache. Data maintenance of a static cache is cost-effective.

- **Cache update**

The alteration of a static cache happens when new members join the organization, and when the subscription network of a FDN changes. A joined new member proposes a set of FDN which includes ported numbers and non-ported numbers to the system administrator. Accordingly, a PBX with OAI (referred as PBX for short hereafter) sends a message to update the profile in the registered telecommunication network, and launches a batch process to query and cache the corresponding routing addresses of the FDN set. The alteration of routing information of a PBX's FDN is notified from the subscription telecommunications network to the registered PBX. The process to update caches is described in the following.

Every time a NP user changes subscription network, the new routing information must be updated from the new subscription network to the NPAC. Service providers contract with NPAC to gain the right to be informed and to obtain the routing information of NP users who change subscription networks. For the consistency of caches and NPDB, a method for data synchronization is required. A PBX is a subscriber of a public telecommunication service provider. Organizations with PBX register to telecommunications service provider for the service of cache synchronization. The service provider keeps profiles of every registered organization, where every profile records the FDN of the registered organization. According to the profiles, when a ported number changes to another subscription network, the subscription network can notify related PBXs to alter caches. The update of cached data can be transferred by IP-network without going through the subscription network of the PBX. Consequently, the telecommunication resource will not be occupied when updating cached data.

We assume users will not change subscription networks too often. For example, 2% or less of ported numbers will be ported out from an operator every day. The effectiveness of a ported number in a new subscription network is postponed for 24 hours to guarantee data consistency of NPAC and the old and new subscription networks.

- **Cache size**

A cache in a PBX keeps the mapping of a FDN set and the corresponding routing addresses. Every data entry in a cache occupies 186 bits, which consists of a 10- to 15-digit telephone numbers which follows ITU-T Recommendation E.164 [24], a 4-octet routing address of the telephone number, a 4-byte field of the latest modification time, and a 2-bit cache management tag, which is used for cache management and update.

For a static cache, FDN of an organization are kept centralized in a cache of a PBX, every FDN occupies a cache entry without duplication. However, enterprises usually have static contact targets, many numbers may duplicate in several FDN sets of different employees. The required cache size decreases when the proportion of common FDN increases. Assume that there are s members in an organization, the FDN quota of each member is k. Suppose the probability the FDN of a member overlap with other's is r in average, $0 \le r < 1$. Let u_i be the FDN set of user i, and U be the universal set of all individual member's FDN, which can be represented as

$$U = \bigcup_{i=1}^{s} u_i - \bigcup_{i=1,\, j=1, i\neq j}^{s} \left(u_i \cap u_j\right) + \bigcup_{i,j,k=1, i\neq j\neq k}^{s} \left(u_i \cap u_j \cap u_k\right)$$

$$- \cdots + (-1)^{s-1}\left(u_1 \cap u_2 \cap \cdots \cap u_s\right) \qquad (1)$$

From (1), the amount of cache entries ms_{cache} in a PBX should be at least

$$ms_{cache} = sk - \binom{s}{2}kr + \binom{s}{3}kr^2 - \cdots + (-1)^{s-1}\binom{s}{s}kr^{s-1}$$

$$= sk + (-1)^{i-1}\sum_{i=2}^{s}\binom{s}{i}kr^{i-2}, \text{ where } 0 \le r \le 1. \qquad (2)$$

When the FDN of every member is scattered that $r = 0$, the minimum amount of cache entries is

$$ms_{cache} = s \times k. \qquad (3)$$

The required minimum cache entries of a static cache are proportioned to r, the overlap proportion of every member's FDN. Consider that an organization with 8000 members adopts static cache policy to implement a cache in PBX. In the condition that each user has 30 common dialed numbers in average, and the FDN duplication rate $r = 0$, the required cache size is less than 5.5 MB; when $r = 0.3$, the cache size is 4 MB; when $r = 0.5$, the cache size is 2.7MB. In an organization with 15000 employees, each member has 100 frequently dialed numbers in average; if $r = 0$, the total cache size is less than 33.3 MB; $r = 0.5$, the cache size reduced to 17 MB.

It is obvious that the size of a cache is small enough to be located in memory. Therefore, the search of cached data takes less than 0.01 msec memory access latency. Since a NPDB query consumes 2 seconds or less to response the routing information, when NPDB queries are substituted by memory accesses, the database searching delay can be reduced remarkably, and the load of NPDB would be alleviated considerably.

3.3 Cost and performance evaluation

Caches diminish the demand for long-term database queries to reduce the data search delay. At the same time, the operation of caches requires the modification of the signaling system of PBX and the alteration of the operation logic of the local exchange. In this section the cost and the benefit of caches in the fix-lined telecommunication NP service are evaluated.

3.3.1 Cost evaluation

The cost of applying caches to PBX consists of two parts: the initialization cost and the operation cost. The initialization cost includes the cost of modifying the signaling system, and the cost to establish caches in PBX. The operation cost includes the cost to modify profiles in the operator network, and the cost to update cached data in PBX, and the cost to query cache for determining and translating a dialed telephone number. Therefore, the cost of applying caches to PBX can be represented as

$$
\begin{aligned}
C_{cache} &= C_{cache\text{-}initialization} + C_{cache\text{-}operation} \\
&= (C_{modify\text{-}signaling\text{-}system} + C_{cache\text{-}establishment}) \\
&\quad + (C_{modify\text{-}profile} + C_{update\text{-}cache} + C_{query\text{-}cache})
\end{aligned}
$$

The initialization cost is an investment before the operation of a communication system. In the initialization stage, a new rule to process special codes will be added to exchanges without altering the existing operation logic. The size of a PBX cache is small enough to be stored in memory. In addition to install memory in PBX as a cache, PBX must be modified to search caches before forwarding a call setup request, and to append the routing address and a special code to a call initiation request when it is a cache hit.

The major consideration of applying PBX cache is the cost of system operation, which varies with the quantity of FDN and the amount of calls. With regards to the property that the variation of organization members is rare, and the contact targets of an organization has locality, the variation of a FDN set in a PBX and a profile of an organization in the subscription network is infrequent. Because the porting frequency of a telecommunication subscriber is restricted, cache update according to the alteration of routing addresses of numbers in the FDN set is infrequent. The update of cached data is an off-line batch process that can be executed in off-time by IP-based networks. The operation cost of PBX-based caches is reasonable.

3.3.2 NP call setup time evaluation

When caches are applied to PBX, every outgoing call brings about a cache query for routing address translation. The result of cache queries could be cache hit (i.e., the cached data represent the correct routing address of the dialed number) or cache miss. In the case of cache hit, the routing address would be translated in PBX and the call would be routed to the called party directly without querying NPAC. In the case of cache miss, NPAC query is necessary to obtain the routing address of the called party. Location miss (i.e., the cached data is obsolete) will not occur because of the policy of postponed activity of new ported numbers.

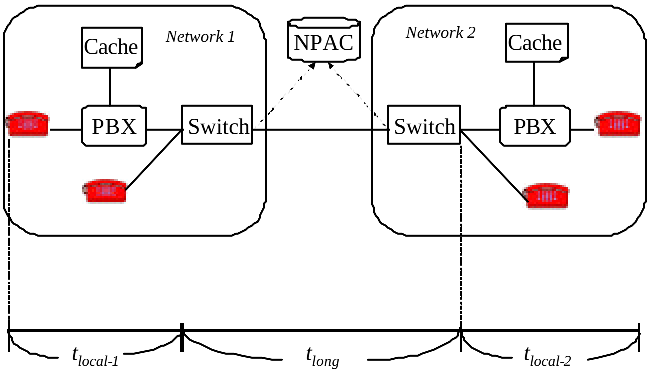

Fig. 3-4 Simplified NP call setup stages and time table

The NP call setup stages can be simplified and represented as Fig. 3-4. Where $t_{local\text{-}1}$ and $t_{local\text{-}2}$ are the local transmission delay in the service region of an operator network, t_{long} is the delay time of cross-operator networks transmission. Let t_{cache} and t_{NPAC} be the routing address translation latency of cache access and NPAC query, respectively. According to Fig. 3-4, the NP call setup time of location-hit of PBX-cache ($t_{cache\text{-}hit}$), cache-miss of PBX-cache ($t_{cache\text{-}miss}$), and conventional scenario (t_{conv}) can be represented as in the following.

- NP call setup time of conventional telecommunication environment:

$$t_{conv} = (t_{local\text{-}1} + t_{local\text{-}2}) + t_{long} + t_{NPAC}$$
$$= t_{local} + t_{long} + t_{NPAC} \qquad (4)$$

- NP call setup time of PBX-cache environment in the case of location hit:

$$t_{cache\text{-}hit} = t_{local} + t_{long} + t_{cache}$$
$$= t_{conv} - (t_{NPAC} - t_{cache})$$

- NP call setup time of PBX-cache environment in the case of cache miss:

$$t_{cache\text{-}miss} = t_{local} + t_{long} + t_{cache} + t_{NPAC}$$
$$= t_{location\text{-}hit} + t_{NPAC}$$
$$= t_{conv} + t_{cache}$$

Let the cache hit rate in a PBX be p, $0= p = 1$, which equals to the probability of that a number in the FDN set being dialed. The call setup time in PBX-cache environment can be depicted as

$$t_{PBX\text{-}cache} = p \times t_{cache\text{-}hit} + (1- p) \times t_{cache\text{-}miss}$$
$$= p \times [t_{conv} - (t_{NPAC} - t_{cache})] + (1- p) \times (t_{conv} + t_{cache})$$
$$= p \times t_{conv} - p \times (t_{NPAC} - t_{cache}) + (1- p) \times (t_{conv} + t_{cache})$$
$$= t_{conv} + t_{cache} + p \times t_{NPAC} \qquad (5)$$

From (5), the call setup time of PBX-cache is proportional to (1- p). The more frequent FDN are dialed, the less the call setup delay is. Comparing the NP call setup time of conventional and PBX-cache applied environment:

$$\frac{t_{PBX-cache}}{t_{conv}} = \frac{(5)}{(4)}$$
$$= \frac{t_{conv} + t_{cache} - p \times t_{NPAC}}{t_{conv}}$$
$$\cong 1 - p \times \frac{t_{NPAC}}{t_{conv}}, \text{ since } t_{cache} << (t_{local}, t_{long}, t_{NPAC}) \qquad (6)$$

From (6), we can say that PBX-cache benefits when p, the probability of dialing a FDN, increases. The figure as shown in Fig. 3-5 depicts the relation of NP call setup time decreases as p increases.

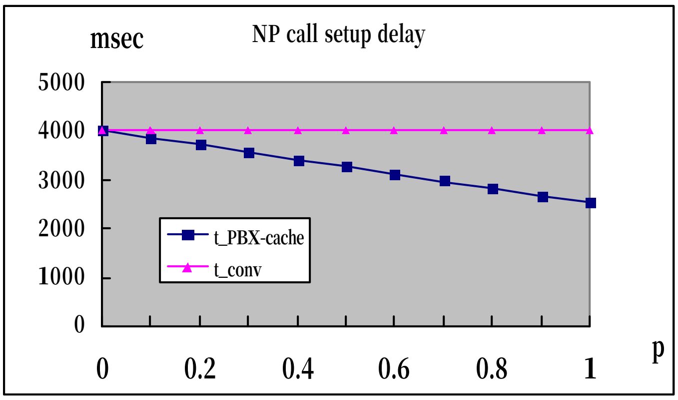

Fig. 3-5 The relation of FDN utility rate (p) and the average call setup time (in msec)

The message transmission time of every trunk can be represented as *(message length / transmission rate of the trunk) + (trunk length/velocity of light or velocity of electrons).*

With respect to Fig. 3-4, the message transmission delay within an operator network (t_{local}) and between operator networks (t_{long}) can be calculated. The time for querying routing information from NPAC (t_{NPAC}) should be less than 2 seconds [5], and the memory access time for searching data form a cache (t_{cache}) is assumed as 0.001 msec.

Assuming the transmission media beyond PBX is optical fiber, and within a PBX service region is twisted pair. The scope of connections within a PBX is 1km, connections between PBX and the local exchange spread from 1km to 10 km. The scope of long-term connections among different telecom service networks follows the geographical length of Taiwan (it is 390km), assuming the longest connection is less than 450 km. The size of calling signals is assumed to be 180 bits (*ISUP IAM* and added special codes). Suppose the time to process an *ISUP IAM* message is 100msec, which includes the time for signal codec, routing process, and network resource allocation. For service providers, each database handles 8 million calls per day; that is, 435 calls per second [23]. NPDB query must complete within 2 seconds or less [15]. A PBX handles 2000 calls per hour, and 8 calls per second in the rush hours. The delay of searching a cached entry is less than 1/100 millisecond in average.

Considering there are 10 thousands members in a PBX service region, every user has a set of 30 FDN in average. The setup time of a call is 4000 milliseconds without implementing caches to PBX. Fig. 3-5 demonstrates the relation of FDN utility rate p and the average call setup time with a cache in PBX. When 30% of the calls are made to frequent contact objects (i.e., 30% of cache queries are hit. While location miss will not happen, all the calls are location hit), the setup time of a call reduced to 3550 msec in average, 11.25% of the call setup time is saved. The NP call setup time decreases when the dialed numbers has locality and the probability of dialing a number in the FDN set increases. When 70% of the dialed numbers are FDN, and the call setup time decreased to 2950 msec. 30% of call setup time is reduced.

Assume the average calling rate is 450 calls per hour of an organization with 10-thousand employees. 70% of the outgoing calls are set to numbers in the FDN set. iNetwork with can shorten 0.7 × (4000-2950) = 735 milliseconds call setup delay for every call in average. Thus, the shortened call setup time is (0.735 second × 450 calls) = 330.75 seconds/per hour. For 8 hours office time per day, the cache can save the organization (330.75 × 8)/60 = 44.1 minutes = 0.735 man-hour per day. For 22 working days per month, (0.735 × 22 × 12) = 194.04 man-hour per year are saved for an organization.

In the peak-time the NPDB workload capacity will not detain the call setup process. As the traffic originated from enterprises increase in the peak time, the increase of FDN utilization prominently alleviates the requirement of NPDB accesses. Assume there are 10-thousand calls per hour arrive a switching center in office hours, and 70% of the calls are dialed from organizations. Let 85% of the 7000 calls are issued from PBX with cache, and the probability of cache hit is p. Consequently, there are (7000 × 85% × p) NP calls are translated without querying NPDB searches. Let 70% of number dialed from organizations are FDN (the cache hit rate is p = 0.7), 4164 calls need not search NPDB; when 90% of calls dialed from organizations are FDN, 5355 NPDB queries are omitted. That is, 41.64% and 53.55% of NPDB workload is alleviated in the case of p = 0.7 and p = 0.9, respectively.

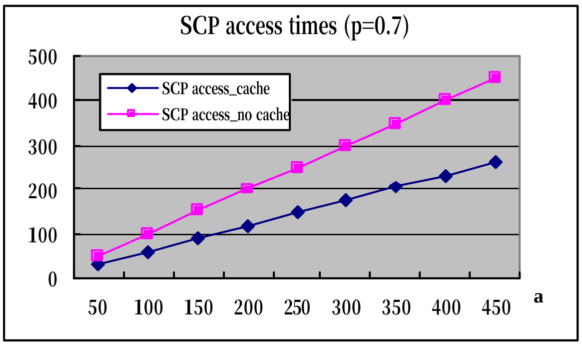

Fig. 3-6 The SCP access frequency

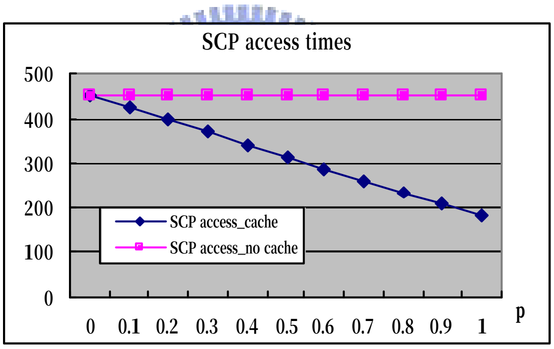

Fig. 3-7 The frequency of SCP accesses when the arrival rate is 450 (calls/hr)

From another point of view, users will hand up a call and re-dial the same number again when the call setup time is too long to be tolerable. The re-dials bring about more signals and consume more network resources. Operators need to take the overhead without any revenue. The shortened NP call setup delay results in the decrease of re-dials, and the network resource can be utilized more effectively.

From the above, applying caches to PBX to enable ported number translation in PBX can remarkably avoid a great part of NPAC queries. Which alleviate the workload of the public switching network and shorten the average NP call setup time. The improved NP call setup efficiency results in better communication resource utilization; therefore, caches on PBX benefit both the telecommunication service providers and users.

Chapter 4. An organization-based cache mechanism for supporting PCS number portability service

4.1 Motivation

Intelligent terminals that can perform ported number translation are not available in mobile telecommunications systems before 2.5G. Therefore, the ported number translation can not be shifted to user terminals to alleviate the workload and traffic load of the public switching network and NPAC.

By studying the mobility and calling behavior of mobile users, we found that a major portion of traffic load is generated from organizations (e.g., enterprises, campuses, government departments) in business hours, and the moving scope of most mobile users is limited. Also, colleagues often need to contact others on the move, many mobile calls are set up to connect organization members in the same service location. Therefore, numbers dialed from a service location which covers an organization will present locality in business hours.

As discussed in chapter 3, shift the effort of ported number translation to organization-based (OGB) networks can alleviate a great part of NPAC queries [26][27]. Therefore, we attempt to apply the notion of PBX-based cache to mobile telecommunication systems to alleviate the traffic and workload of the mobile core network. However, a network entity analogue to a PBX in mobile telecommunications network is not generally available.

In order to utilizing the dialed-number locality of OGB communication, an OGB mobile telecommunications system or component is necessary. A mobile PBX should possess convenient cost-free or low-cost inter-organization mobile telecommunications between organization members, and guarantee global mobile communication service beyond an organization. A local mobile communication system such as IEEE 802.11 based WLAN provides cost-free communication service within an organization, but the restriction of limited communication scope can not provide communication service out of the scope. Organization members have to change handsets and phone numbers when they move off the organization. By contrast, a global mobile communication system such as GSM enables global communication service to organization members without the limitation of communication scope. But the cost is too high to be adopted as an OGB communication system.

An OGB mobile telecommunications system should consist of radio base station and local switching center to provide mobile communication service, the mobile calls among organization members can be processed and routed locally, and the communication cost will be decreased whereas no communication resources beyond the organization are required. The OGB telecommunications system must be able to determine inter-organization from external calls; moreover, it must cooperate with global communication systems to guarantee connections beyond the scope of the organization. When a call generated from an OGB system is terminated to an external user, it should be routed to a global communication system (e.g., GSM, PSTN) according to the prefix of the dialed number (Fig. 4-1).

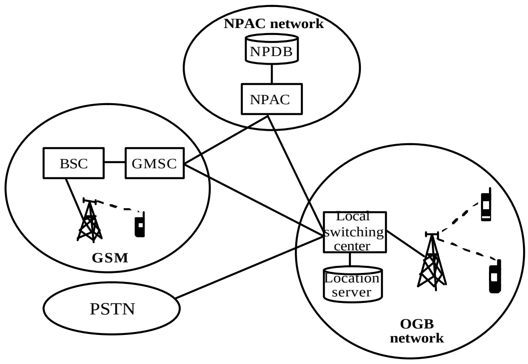

Fig. 4-1: The concept of an organization-based communication system

For the purpose, based on Internet Protocol (IP) we develop an organization-based mobile telecommunications system—iNetwork—which fulfills the requirement of cost-free communication within an organization, and supports GSM-compliant mobile service that guarantees global mobile communication when users leave the service region of an organization [28]. Since frequent dialed numbers and their corresponding routing information can be stored in the memory of iNetwork, the iNetwork can effectively enhance the efficiency of mobile NP service.

4.2 Design of iNetwork

For the need to provide low-cost global mobile telecommunications service, iNetwork is developed as an IP-based OGB mobile communication system, which is implemented as IP-based software components. iNetwork is designed as GSM-compatible to enable mobile communication service within and beyond an organization. GSM BTS (Base Transceiver Station) is adopted as wireless access point. Users utilize GSM mobile handsets or personal computers with GSM access interface as communication terminals. The mobile service of iNetwork follows GSM system because it is the most prevalent mobile telecommunications system in the world, and it possesses the capability of data services.

Because iNetwork is the asset of an organization, only iNetwork subscribers are allowed to access iNetwork. iNetwork users must be authorized to access iNetwork service. The authentication process takes place when a user moves into an iNetwork service region. Within an iNetwork service region, the communication of organization members is processed and routed to the called party via the local IP network; intra-organization communications are cost-free just like using a PBX.

A large-scale organization or enterprise may have several subsidiaries distributed over different physical regions. To provide communication service between members in different subsidiaries, several iNetworks can be organized as a community to share resource and to enable intra-organization communication. Similarly, several iNetworks of different organizations can form an alliance or community to enlarge the service region of iNetworks. The communication between different iNetworks of a community is IP-based. Just like intra-organization communication, the cost of intra-community communications can be reduced or free or charge for organizations and members.

In addition, iNetwork users have the demand to communication with users of other communication systems (e.g., GSM, PSTN). By contract, iNetwork registers to GSM as an add-on service and shares the numbering plan of the subscription GSM operator. Therefore, the subscription GSM can determine iNetwork subscribers by users' GSM IMSI (International Mobile Subscriber Identify – the subscribers' identity numbers) and MSISDN (Mobile Station ISDN – the subscribers' phone numbers). When one of the call-leg is out of iNetwork service region, the connections will be established via the subscription GSM. When moving off iNetwork service regions, iNetwork subscribers register to the visiting GSM network to keep connection. Thus, iNetwork users can obtain GSM services without losing connection when they move out of iNetwork service region.

The routing and roaming information of subscribers is the most significant information to switch and route traffic between iNetwork and GSM. Both iNetwork and GSM should provide a mechanism and interface for information exchange.

4.2.1 iNetwork architecture

Fig. 4-2 presents the iNetwork architecture. Except the off-the-shelf GSM BTS, iNetwork components are implemented as software components connected to the IP-based network to perform IP telephony and GSM-compliant mobile telecommunications service [28].

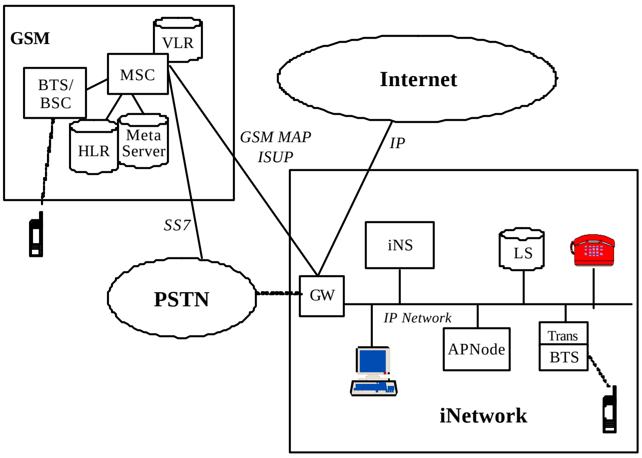

Fig. 4-2 iNetwork architecture

As depicted in Fig. 4-2, every iNetwork service area consists of at least an iNS to provide switching and routing functions; a LS to maintain user and location information; one or more APNodes to handle requests from users, to collect user information, and to manage network resources; several BTS to support radio access to iNetwork users, and they are bound with Trans that translate protocols between the IP network and GSM system. iNetwork components are connected with IP-based network. Every iNetwork has to register to iNetwork system to obtain the system access right. The communication model of iNetwork components is shown in Fig. 4-3, iNetwork provides GSM MAP, ISUP, and SIP interface to connect with GSM network. SIP (Session Initiation Protocol) [29] is adopted as the signaling protocol of iNetwork system administration, session control, and call handling [30]. iNetwork components and their functions are described in the following.

- *iNS*: iNS is the primary component which provides switching and routing functions. The interworking protocols ISUP, MAP, and SIP are adapted and implemented in iNS as shown in Fig. 4-3. iNS monitors and processes signals of the surrounding APNodes to provide calling service. Accounting information of the source and destination addresses can be recorded by iNS.
- *LS* (*Location Server*): LS is a data repository that is responsible for maintaining user and location information. LS acts as the functions of HLR and VLR in an iNetwork. In addition, considering the need to connect subsidiaries of an organization as a telecommunications community, the organization members belong to different subsidiaries/iNetworks should be able to access the communication of the entire iNetwork community. Therefore, a LS includes three parts: *home-LS*, *community-LS*, and *acting-LS*.
  - The Home-LS part maintains the information of an iNetwork's subscribers. Every entry in the home-LS part must keeps MSISDN and IMSI of a user as

the ID, the routing address of the current visiting network (including GSM and iNetwork), and the state of the user (active/non-active).

- The community-LS part contains the information of the subscribers of other iNetworks of the community. Every entry in this part consists of MSISDN as user ID, the user's subscription iNetwork, and the routing address of the user (in the format of *user@subscription_iNetwork*).
- The acting-LS part maintains the information of iNetwork users who reside and are active in the iNetwork. Every entry in the acting-LS consists of MSISDN as the user ID, the user's subscription iNetwork, a temporary ID (as an TMSI) as a ticket to access iNetwork service, and the routing address the user (*user@this_iNetwork*).

Every iNetwork user has a corresponding entry in the LS of his/her home iNetwork and all the other iNetworks of the community. When roaming to an iNetwork of the community, an entry will be added to the active-part of the visiting network's LS. The interactions of LS and iNS perform analogue service functions as GSM HLR and VLR.

- *MetaServer*: MetaServer is a conceptual component that provides location information for iNetwork and GSM users. For security concerns, MetaServer resides in GSM network. It is a GSM component that performs the query interface of HLR. Whenever an iNetwork user moves to and registers to GSM, user and related location information can be retrieved via MetaServer. iNetwork system queries MetaServer to obtain information of roaming users.
- *APNode* (*Application Node*): It is used to handle the requests from users, to collect user information, and to allocate and administrate network resources. APNode filters and forwards requests to iNS. APNode informs iNS of user requests. One or more APNodes connects to an iNS by dedicated link, the iNS administrates the APNodes to perform communication services. The extended functions and services can be implemented in APNode.
- *BTS (Base Transition Station)*: This is the component that implements GSM BTS. BTS consists of RF transmitters, receivers, and signaling equipment that performs radio access interface to GSM-complied terminals. BTS enables iNetwork to provide communication services wirelessly.
- *Trans (Translator)*: Trans is a protocols translator that is bound with BTS as a gateway between IP-based network and PSTN.
- *Gateway*: Gateway provides translation functions to translate protocol format and communication procedure among Mobile Application Part (MAP), Signaling System No. 7 (SS7), and Internet protocols. Gateways in iNetwork support routing functions of signaling and traffic flows.

- *Terminals*: iNetwork terminals can be personal computer hosts or GSM-compliant terminals.

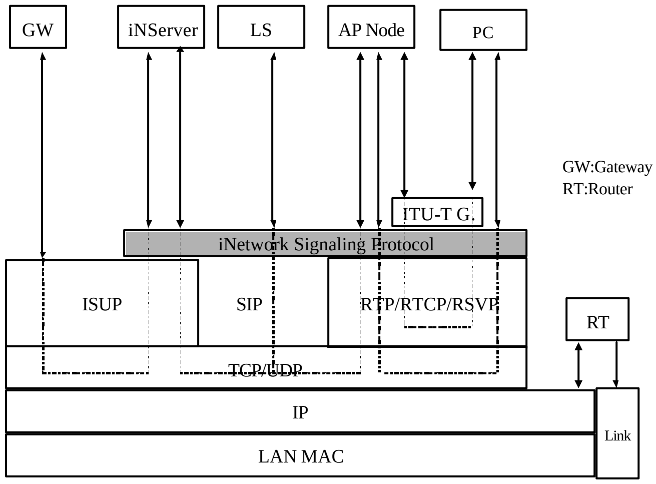

Fig. 4-3 iNetwork signaling protocol stack

Registration requests may come from mobile terminals or from computer hosts (Fig. 4-4). Mobile users access iNetwork system via BTS/Trans—the radio access point of iNetwork system, all the messages and requests are forwarded to the APNode connected with the BTS/Trans. Requests from computer hosts are broadcasted through the enterprise network and waits for the responses from corresponding APNode of the area. APNodes allocate and control network resources for users and communication sessions. For authentication and authorization, APNodes assign each user with a transport address and an access key to obtain system services. Requests are forwarded from APNode to iNS. iNS processes requests of register, call setup and tear down, and other teleservices. With embedded proxy and redirection functions, iNS can handle the switching and routing processes to handle communication services.

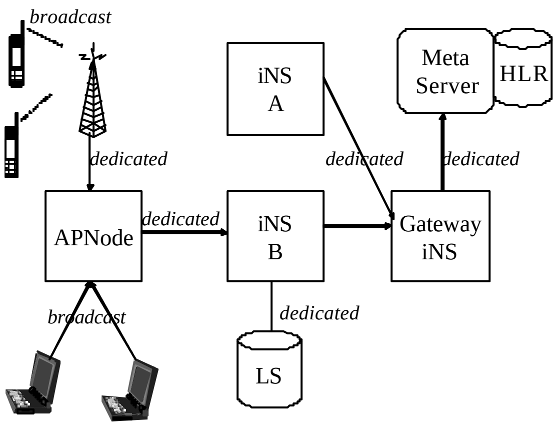

Fig. 4-4 iNetwork Register.

Interoperations between other communication systems (including other iNetworks) are performed on iNS. Different iNetwork systems share user and location related information to enable wide area communication services. The interoperations between iNetwork and GSM to exchange users and location information are achieved by MetaServer queries. Thus iNetwork can provide worldwide communication services without the limitation of IP-based networks service range. Interoperations between MetaServer and iNetworks follow the GSM MAP and ISUP to enable call processing, call forwarding, and other teleservices [35].

To enable the intercommunication between iNetwork and GSM, the addressing method must integrate the communication models of the IP network and GSM. The email-like addressing mechanism (i.e., *user@host*) forms a common addressing space for the intercommunication of GSM, PSTN, and the IP network. To consist with ITU-T E.164 [23] and the existed switching systems, iNetwork addresses follow IETF electronic numbering (ENUM) [31], which can be used to identify a switch, an end-point, or an individual subscriber. The *user* portion is a decimal digit string that indicates a subscriber, *host* indicates an IP network address; aliases are allowed for easy memorizing. In order to comply with the numbering plan of SCN (switching circuit networks), the subscriber number must be a string of 10-15 digits.

4.2.2 iNetwork operation model

As an organization-based telecommunication system, iNetwork acts as a mobile PBX to provide cost-free intra-organization telecommunication service. When an organization consists of several subsidiaries, several iNetworks may cooperate as a communication community to enlarge the service region. iNetworks of the community share communication resources, and the subscribers of each iNetwork can obtain communication service in every service region of the community. When organization members move off the service region of iNetworks, they register to contracted or roaming-agreed GSM operators automatically to obtain communication services. On the other hand, iNetwork users have the demand to communicate with users of other telecommunication systems. Therefore, an iNetwork cooperates with GSM operators to complement the communication service beyond iNetwork service areas and to enable global telecommunications. The cooperation also enhances the loyalty of an organization and its members to the global service providers.

To provide intra-organization, intra-iNetwork community, and inter systems telecommunication services, an iNetwork needs to distinguish organization or community members from others. For the purpose, an iNetwork must keep the information of registered users (including subscribers of the iNetwork and members of the community) in LS. In the process of call origination, the iNS consults LS by caller's IMSI for authentication. If the caller is not authenticated, the call is bypassed or routed to the subscription GSM network. After that, the iNS consults LS by the callee's MSISDN to determine the identify and the location of the called party. According to the network system the called party resides, iNS routes the call to the callee via IP network directly or via the subscription GSM network.

The cooperation models of intra-iNetwork, intra-community, and inter communication systems are illustrated in the following.

■ **Intra-iNetwork communication**

Because the LS of an iNetwork keeps the information of all the subscribers of the iNetwork and the iNetwork community, an iNetwork has the ability to determine that whether the called party is an iNetwork subscriber, and whether the callee resides in the local service region or not.

The home- part of an iNetwork's LS maintains the visiting network of all the home subscribers, and the community-part of the LS maintains the home networks of other community users. When a request of call-origination is issued, an iNetwork consults the home- and community-part of the local LS in turn to determine whether the called party is an iNetwork user (Fig. 4-5 setup 1 and 2 in white circle). When the call is set to an Network user, the origination iNetwork will consults acting-part of the local LS to determine whether the callee resides in the local service region (Fig. 4-5 step 3 in white circle). When the caller and callee resides both reside in the same iNetwork service region, the call is routed and transmitted directly via IP network without requiring the communication resource beyond the scope of the iNetwork [27].

Otherwise, the call will be routed to the visiting network (by the information in home-part) or to the home iNetwork (by the information of community-part) accordingly.

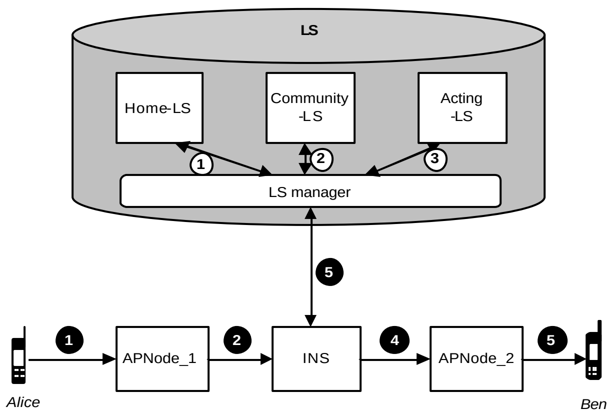

Fig. 4-5 Intra-iNetwork communication

■ **Intra-iNetwork community communication**

Several iNetworks may cooperate as a communication community to share communication resources and to enlarge the service region. Members of the community can obtain communication service in all iNetworks of the community. The authentication process is triggered when users move from an iNetwork to another, or move in the service region of an iNetwork. The state and location information of subscribers are informed to iNetwork by register process. Based on the authentication and location information of subscribers, iNetworks can provide roaming, call handling, and handoff services to users.

The process of subscriber location register is shown in Fig. 4-6. GSM_1 is the subscription GSM of iNetwork_B. User Alice subscribes to iNetwork_B, she is also a subscriber of GSM_1. Whenever Alice enters the service region of a visiting iNetwork, the latest location information is updated to the subscription iNetwork_B by a *MAP_UPDATE_LOCATION_AREA* message. The register information consists of the IMSI and the MSRN (Mobile Station Roaming Number) of Alice. iNetwork_B keeps the location and routing information in local LS, and forwards the register information to the subscription GSM_1. By way of register process, the location information of Alice is kept in both the subscription iNetwork_B and the subscription GSM_1.

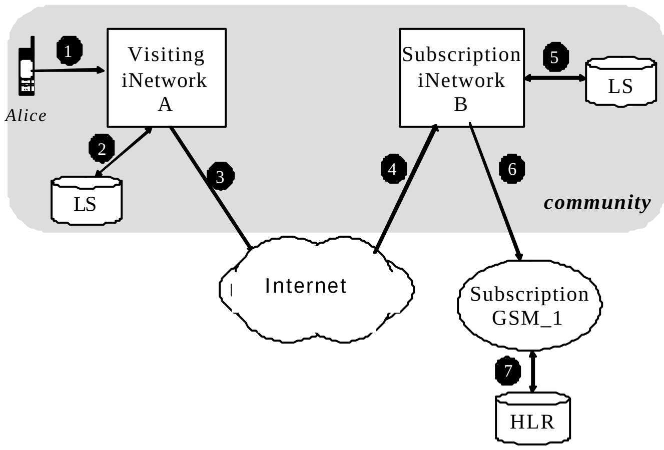

Fig. 4-6: iNetwork register process

Assume IP connections are available among iNetworks of a community. Communications of iNetwork users of the same community is basically transferred through the IP network. The call setup process in an iNetwork community is illustrated in Fig. 4-7. The origination iNetwork_A receives and determines a request to set up a call to an iNetwork user. When the origination iNetwork queries LS and confirms that the called party resides in the local service region, iNetwork_A forwards the call to the called party directly without requiring resources beyond iNetwork_A. If the called party did not reside in the local service region, the routing address of the callee's subscription iNetwork is appended to the dialed number (i.e., Alice@iNetwork_B), and a *SIP INVITE* message is issued to the callee's subscription iNetwork_B. iNetwork_B receives the message and queries LS for the routing address (a MSRN, which indicates the address of the termination switch which the MS attaches) of the callee, then routes the request to the termination iNetwork_C to set up the call.

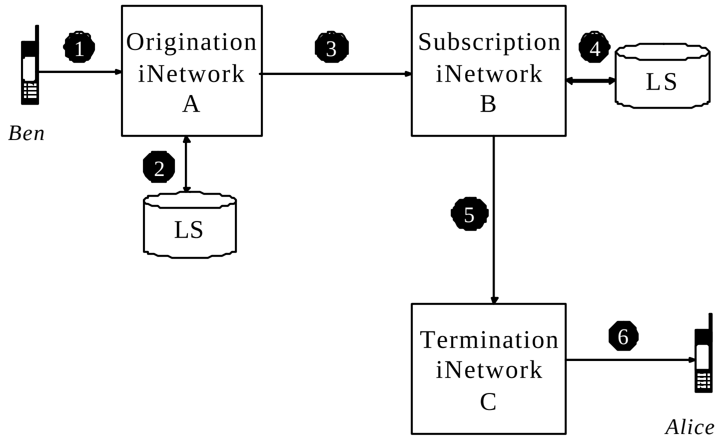

Fig. 4-7: Call setup in an iNetwork community

■ **Intercommunication of iNetwork and GSM**

To complement the connection beyond the service region of an iNetwork or an iNetwork community, iNetwork cooperates with GSM as an add-on service or a registered business subscriber. All registered iNetwork users are also GSM subscribers that can access the communication resources of the GSM network. iNetwork subscribers register to the GSM network when moving off iNetwork service regions. The location information of an iNetwork subscriber is shared between the subscriber's subscription iNetwork and the subscription GSM.

The interoperation between iNetwork and the subscription GSM follows the SS7 (Signaling System 7) signaling and communication model. The protocol for exchanging call handling messages is ISUP, routing information exchanges are based on GSM MAP (GSM Mobile Application Part), and the routing of calling signals is conveyed by TCAP (Transaction Capabilities Application Part) commands. To solve the problem of different naming plans of the IP-based communication network and GSM, a new addressing policy following IETF electronic numbering (ENUM) [31] is recommended. ENUM provides a standard process by which a full-length E.164 telephone number (e.g., +886-3-5712121) can be converted into a domain name (e.g., 1.2.1.2.1.7.5.3.6.8.8). This process requires the renewal of conventional switching networks.

When a cooperation contract of iNetwork and GSM is established, the numbering plan of iNetwork is included to that of the GSM. Thus the GSM can recognize iNetwork numbers. Every iNetwork subscriber has an iNetwork address in the format of *MSISDN@network*. When an iNetwork subscriber leaves the service region of iNetworks and enters the service area of a GSM service provider, he/she registers to GSM with his MSISDN. The visiting information will be sent from the visiting GSM to the subscription GSM. GSM filters iNetwork subscribers according to the numbering plan and updates the latest location of roaming iNetwork subscribers to the subscription iNetwork via MetaServer.

Fig. 4-8 illustrates the process of setting a call from iNetwork service region to a roaming iNetwork subscriber. Ben in an iNetwork service region initiates a call to Alice by dialing Alice's MSISDN. The origination iNetwork-A queries LS to determine the called party does not reside in the local service region, and routes the request to the subscription iNetwork-B of Alice by the MSISDN. iNetwork-B queries LS for the roaming information of Alice and routes the request to the termination network. If the visiting GSM is the subscription GSM of iNetwork-B, iNetwork-B forwards the request to the subscription GSM to set up the call. Otherwise, iNetwork-B forwards the request to the subscription GSM, then the subscription GSM routes the request to the termination GSM to set up the call.

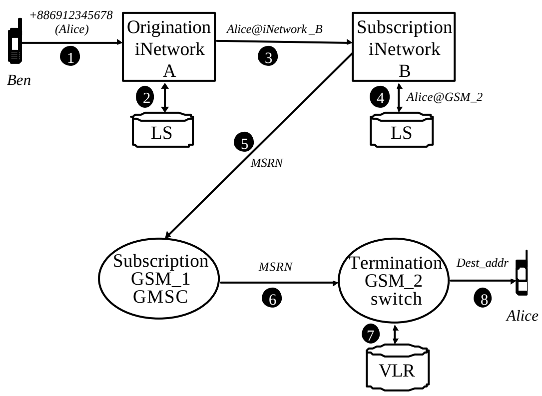

Fig. 4-8 Set up a call to a roaming iNetwork user

4.3 Providing mobile NP service by utilizing OGB mobile telecommunication system

According to the discussion in chapter 3, we found that the outgoing calls generated from an organization have dialed-number locality. Keeping the routing information of the frequently dialed numbers (FDN) in OGB communication system will enable the ported number translation in the early stage of NP call setup without querying NPAC. Thus, a large part of the workload of ported number translations will be shifted to the OGB network, and the workload of NPAC and the traffic load of the public switching network will be alleviated to improve the efficiency of NP call setup.

When iNetwork is adopted as the local mobile communication system of an organization, iNetwork interrogates GSM HLR for the routing information of ported numbers. Based on the property of dialed number locality of OGB network, the obtained data is cached in iNetwork to prevent accessing the same data reiteratively. Before setting up an outgoing call, iNetwork queries local cache for the subscription network of a dialed number before querying NPAC. When the routing information of the dialed number is kept in the local cache, the call can be routed to the subscription network of the dialed number directly without time-consuming NPAC query and the related message processes.

A cache of an OGB network is used to reach the subscription networks of the most frequently dialed numbers directly without querying NPAC for ported number translation. NP users may also be iNetwork subscribers. Calls between iNetwork users who reside in iNetwork service regions are set via iNetwork communication resource. But when iNetwork users move off service regions of the iNetwork community, they are active in GSM networks. Consequently, OGB cache must keep the public subscription networks of FDN, even if the FDN belong to iNetwork subscribers. Every entry in a cache consists of a telephone number, the subscription network of the number, and the routing information to reach the subscription network of the number. Also temporal data is kept in the cache entry for cache management.

The processes for iNetwork users to set up calls to NP or non-NP users in iNetwork or in other network systems are depicted in the following.

4.3.1 Call setup process in an iNetwork

The call initiated from iNetwork users can be terminated to a user in the same iNetwork or in the same iNetwork community, also it can be terminated to a GSM user. An iNetwork needs to identify whether the called party locates in iNetwork or in other communication system.

When a call is set to a user in the service region of iNetwork, the call should be routed to the user via IP-based network directly without occupying any resource of the public telecommunication system. Hence, when an iNetwork receive a request of call-origination, it consults the home- and community-parts of the local LS to determine whether the call is set to an iNetwork subscriber. If the called party is an iNetwork subscriber, the iNetwork routes the call to the called party's visiting iNetwork. The visiting iNetwork of the called party consults the acting-part of its LS for the called party's destination address to set up the call.

Fig. 4-9 illustrates the call setup process between two iNetwork subscribers in the same iNetwork community. The origination iNetwork-A receives a call initiation request (step 1) and determines the dialed number indicates an iNetwork user. iNetwork-A queries home- and community-part of the local LS first to determine if the called party resides in the local service region or in other iNetwork of the community (step 2). When the called party resides in other iNetwork, iNetwork-A routes the request to the callee's subscription network by the obtained routing information (step3). The callee's subscription network queries LS for the routing information of the termination network (step 4), and routes the request to the termination network (step 5). The termination network pages the called party's handset and forwards the request to the called party to set up the call (step 6).

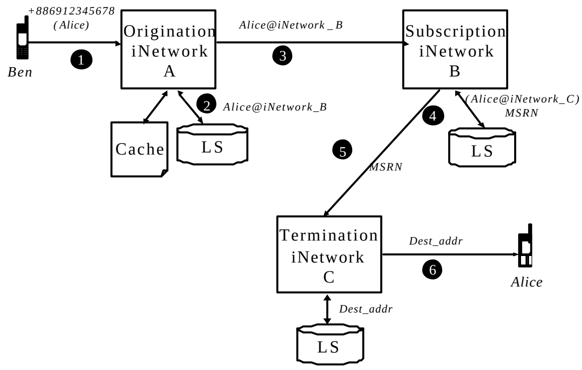

Fig. 4-9: Set up a call in an iNetwork community

The call setup process is depicted as shown in Fig. 4-10.

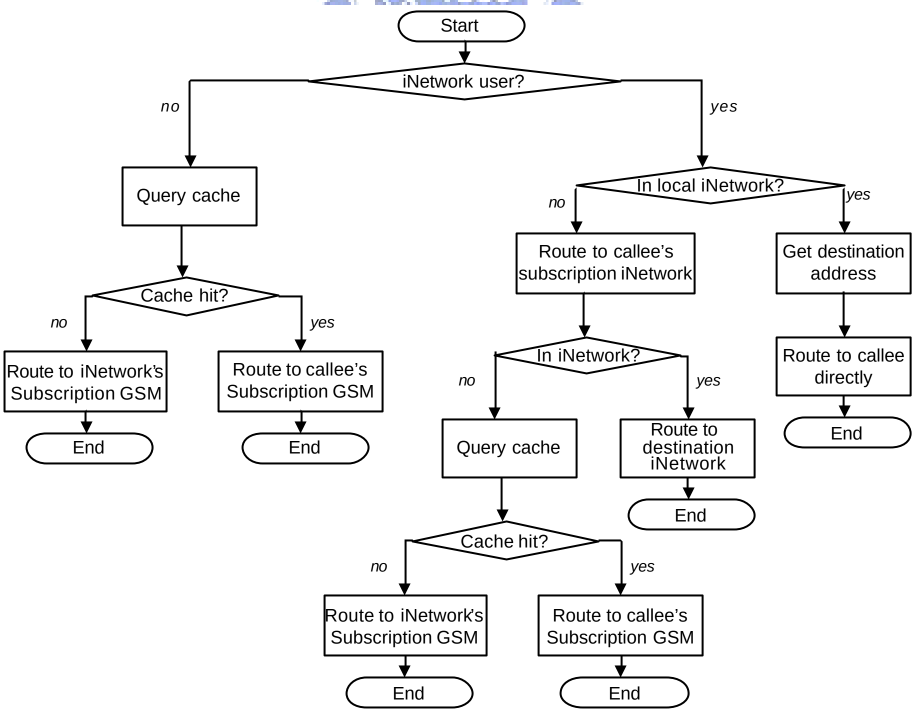

Fig. 4-10 Call setup process of iNetwork

4.3.2 Set up an NP call to the GSM system

If the called party is not an iNetwork subscriber, or is a roaming iNetwork subscriber who moves off iNetwork service regions, the call will be routed to the public telecommunications network. iNetwork caches keeps the subscription networks of FDN that can be utilized to perform ported number translation in organization network. Before routing an outgoing call to the public telecommunications network, iNetwork consults local cache for the subscription network of the dialed number. In the case of cache hit, the dialed number is a frequently dialed ported number that the routing information can be confirmed without NPAC queries.

Every iNetwork must have a direct connection with its subscription GSM network, but the connection with other GSM networks is not necessarily available. The communication between an iNetwork and the non-subscription GSM networks is often supported with the help of the subscription GSM network of the iNetwork.

Fig. 4-11 elaborates the process of setting a call from iNetwork to GSM network. The origination iNetwork-A receives and determines a request to setup an outgoing call (step 1). Before routing the request to its subscription GSM network, iNetwork-A consults local cache for the subscription network of the called party (step 2). If it is a cache miss, iNetwork-A forwards the request to its subscription GSM_1 network (step 3 in black circle). GSM_1 process the call by the conventional call-origination process (step 4 to 9 in black circle).

In the case of cache hit, the dialed number is a FDN and the routing information of the callee's subscription network is determined. iNetwork forwards the request of call-origination with the callee's routing information to the subscription GSM_1 (step 3 in white circle). GSM_1 routes the call to the subscription network of the called party directly without the process of ported number determination and ported number translation (step 4 in white circle). The subscription network of the called party processes the call by the conventional call-origination process (step 6 to 9 in black circle).

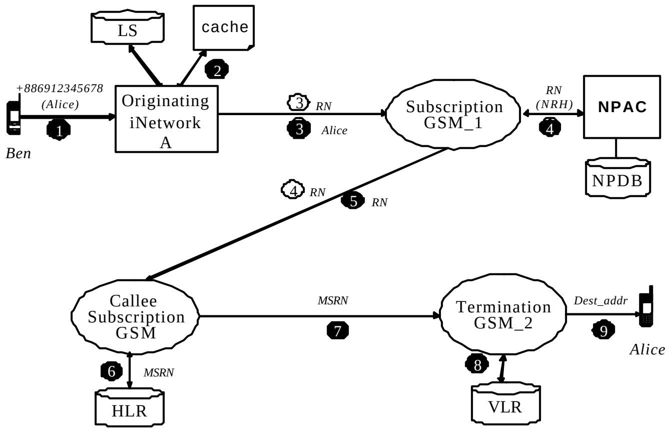

Fig. 4-11: NP call setup between iNetwork and GSM

4.3.3 Side effects of OGB-based cache

■ **Miss routing**

The worse case of NP call setup occurs when mis-routing happens. Mis-routing happens when the cached routing information is obsolete. The call will be routed to a wrong destination and results in the fail of call setup process.

As illustrated in section 4.4, iNetwork caches are implemented as static cache that the update of routing information is a batch process rather than a real-time process executed whenever a number in the FDN set is ported. When a NP subscriber changes service provider, the effectiveness of the change is postponed for couple hours to guarantee the routing information kept in NPDB is consistent with that in the old and new subscription networks. The update of cache is executed in downtime, and accessed to the cache is forbidden during the update of cache. Hence, the case of mis-routing can be prevented.

■ **Handoff overhead**

The service region of an OGB network is limited that may increase the amount of handoffs during a call. However, the moving speed of users in an organization is not fast, such that we assume that the calling time in a service region is long enough to cover the cost of handoff overhead.

4.4 Cache establishment and update

Organization-based caches can effectively reduce the requirement of database queries and enhance the efficiency of the NP service. However, the policy of cache establishment and the complexity of cache management influence the cost and the benefit of caches.

The data be cached should be as many as possible to enhance the cache hit rate, but the size of a cache is restricted that the amount of data entries in a cache is limited. In order to improve the hit rate of caches to enhance the efficiency of the call setup process, caches should be able to expose the communication habits of users of the service region.

The nature of mobility broadens the moving scope of mobile users. From another point of view, the amount of users a single service region serves increases with the moving in and out of users. But in an organization mobile network, the alterations of members and the frequently dialed number (FDN) of users are infrequent. The moving scope of organization members in business hours is usually limited in the coverage of the organization communication system, and the amount of users moving in and out of the service region is limited that a cache with restricted size can perform appropriate hit rate. For a static cache has the advantage of saving management effort to maintain the cached data, as described in section 3.1.1, we adopt static cache policy to cache the regular or frequently dialed numbers of users for a long-range observation. The measurement of cache size is analogous to that in section 3.1.1.

Usually the frequently dialed numbers (FDN) represent the cooperative organizations, providers and suppliers, agents, etc. of the organization, also include the families and friends of organization members. The establishment and alteration of a static cache is manually performed by a system administrator. When a member joined an organization, the member proposed a set of FDN to the system administrator. The system administrator interrogates the contracted GSM (the subscription GSM) to obtain the corresponding routing information of the FDN set.

When a NP subscriber changes service provider, the effectiveness of the change is postponed for couple hours to guarantee the routing information kept in NPDB is consistent with that in the old and new subscription networks. The subscription network maintains profiles to keep the FDN sets of registered iNetworks. When a number in the FDN set is ported to another service provider, the organization's subscription network issues a notification message to the registered iNetwork according to the profile to notify the alteration of routing information. The interrogation is a batch process and is transmitted via IP network without occupying telecommunications resource.

4.5 Cost and benefit analysis

Bandwidth is occupied during the process of call setup. Additional message processing and transmission is necessary and inevitable in operator network for providing mobile NP service, and the long delay time for setting up NP calls brings about more bandwidth consumption for the operator network. For saving bandwidth for the communication system and economizing the communication resources for additional message processing and transmission, the call setup time should be shortened. A shorter call setup time also prevent users from long waiting for putting through the connection.

Apply caches in OGB network can prevent a lot of NPAC queries, and the average NP call setup delay can be shortened according to the prevented NPAC queries and the alleviated workload and traffic load of the operator network. The evaluation is analogous to that illustrated in chapter 3. In this section, we only investigate the NP call setup time in different routing cases and the cost-benefit of OGB network.

4.5.1 Comparison of NP call setup time

The time required to initiate a call (denoted as t_c) can be represented as

t_c =*transmission delay + data query delay + service delay.*

Transmission delay includes IP-based network transmission delay (t_{IP}), that is the time to transmit messages in one or more iNetwork service areas, and GSM transmission delay (t_{GSM}), which is the time to transmit messages through one ore more GSM networks. Data query delay includes the latency of querying data form NPDB (t_{NPDB}), from HLR/VLR ($t_{HLR/VLR}$), and from local cache/LS ($t_{cache/LS}$). The size of a local cache and LS is small that can be arranged in memory. The query delay $t_{cache/LS}$ is much smaller than $t_{HLR/VLR}$ and t_{NPDB}. The time to query NPDB data should be less than 2 sec [15], the time for local database query should be less than 350 msec [23]. Consequently, $t_{cache/LS} << t_{HLR/VLR} < t_{NPDB}$. The service delay ($t_s$) includes of the delay time for user authentication, for codec, and for message processing.

The transmission delay varies according to the different routing paths and adapted transmission networks. In order to describe the transmission delay precisely, IP-based transmission delay t_{IP} is divided into $t_{IP\text{-}local}$ and $t_{IP\text{-}long}$ to represent the different routing scenarios of intra-iNetwork and intra-community transmission.

The call setup time with respect to different call setup scenarios are presented as in the following:

- Setting a GSM non-NP call in the conventional mobile telecommunications environment. When receiving the call-origination request, the call origination network will determine the dialed number is not a ported number, and routes the request to the subscription network of the called party. The called party's subscription network consults HLR for the routing information of the user's visiting network, and routes the request to the visiting network. The visiting network consults VLR for the destination address of the called party to set up the call. Thus, the call setup time of this scenario is represented as

$$t_{conv} = t_{GSM} + 2\times t_{HLR/VLR} + t_s \qquad (1)$$

- Setting a GSM NP call in the conventional mobile telecommunications environment. The call origination network needs to query NPAC for the subscription network of the called party, and routes the request to the called party's subscription network according to the obtained routing information. Then, the called party's subscription network consults VLR for the destination address of the called party to set up the call. The call setup time of this scenario is represented as

$$\begin{aligned} t_{NP} &= t_{GSM} + t_{NPDB} + 2\times t_{HLR/VLR} + t_s \\ &= t_{conv} + t_{NPDB} \\ &= (1) + t_{NPDB} \qquad (2) \end{aligned}$$

- Setting an intra-iNetwork call when iNetwork is adopted as a mobile OGB telecommunication system. The call origination iNetwork queries local LS and finds the calling and called party both reside in the local iNetwork service region. The call is routed to the called party directly via the local IP-based network without consuming communication resource beyond the iNetwork (Fig. 4-5).

$$t_{iN} = t_{IP\text{-}local} + t_{cache/LS} + t_s \qquad (3)$$

- Setting an intra-community call between users who reside in different iNetworks of the same communication community. The call origination iNetwork queries local LS and finds the called party is a community member who does not reside in the local iNetwork service region. The call origination iNetwork obtains the routing address (in the form of *user@network*) of the callee's subscription iNetwork to route the request. The subscription iNetwork consults local LS for the routing information of the called party and routes the call to the termination network. The termination network consults local LS for the destination address of the called party to set up the call (Fig. 4-7). The call setup time is

$$t_{iN\text{-}community} = t_{IP\text{-}local} + t_{IP\text{-}long} + 3\times t_{cache/LS} + t_s \qquad (4)$$

- Setting a call to a GSM NP user, in the case of cache hit. The dialed number is a FDN and the routing information of the dialed number is cached. The routing information will be attached in the call-origination request when the request is routed to the subscription GSM network of the call origination iNetwork. The subscription GSM of the iNetwork routes the request to the callee's subscription network directly without triggering ported number translation process (Fig. 4-11).

$$\begin{aligned} t_{iN\text{-}GSM\text{-}hit} &= t_{GSM} + t_{cache/LS} + t_{HLR/VLR} + t_s \\ &= t_{NP} - t_{NPDB} + t_{cache/LS} \\ &\cong (2) - t_{NPDB} \end{aligned}$$

= (1) (5)

- Setting a call to a GSM NP user, in the case of cache miss. The routing information of the dialed number is not cached. Hence, iNetwork routes the call to its subscription GSM network, and the conventional GSM NP call process is triggered.

$$t_{iN\text{-}GSM\text{-}miss} = t_{GSM} + t_{cache/LS} + t_{NPDB} + t_{HLR/VLR} + t_s$$
$$= t_{NP} + t_{cache/LS}$$
$$\cong (2) \qquad (6)$$

The access latency of $t_{cache/LS}$ is very small, such that the values of t_{iN} and $t_{iN\text{-}community}$ are close, and the values of $t_{iN\text{-}GSM\text{-}miss}$ and t_{NP} are almost equal. From equations (1) and (5), it is found that when a ported number is a FDN kept in an iNetwork cache, the NP call setup time is almost equal to the setup time of a conventional GSM non-NP call. In equations (2) and (6), when a ported number is not kept in an iNetwork cache, the NP call setup time is almost equal to the conventional GSM NP call setup time.

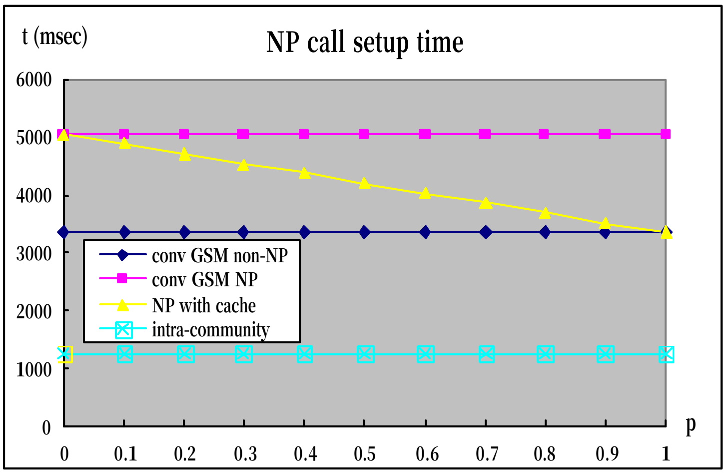

Fig. 4-12 The time for setting up a call from iNetwork to a roaming GSM user

Fig. 4-12 compares the call setup delay of NP call setup in different scenarios. Let *p* represent the probability of iNetwork cache hit. Query routing information from NPAC to translate a ported number takes up a large portion of call setup time. The NP call setup delay decreases with the increase of the probability that FDN are dialed. Communication bandwidth is saved according to the shortened call setup time, and the amount of NPDB queries decreases. In an organization with 10000 employees, the average calling rate is 450 per hour. Assume 70% of the calls are connected to frequently contact targets, that is, 70% of the dialed numbers are cached. The NPDB access both cost 1.5 seconds or more. For 8 hours office time per day, the cache saves the organization 1.05 man-hour cost per day. There are 22 working days per month, the cache will save 277.2 man-hour per year for an organization.

8 hours ×450 calls per hour ×0.7 FDN utilization ×1.5 sec = 1.05 hours

The time to set up a call within an iNetwork community is less then 1500 msec. Comparing with conventional GSM call setup time, every intra-community call can save more than 2 seconds for call setup latency. If 70% of the calls generated from organization members are terminated to other organization members in the same iNetwork community, the call setup time be saved is 1.4 hours per day.

8 hours ×450 calls per hour ×0.7iNetwork utilization ×2 sec = 1.4 hours

The shortened NP call setup delay also benefits in reducing re-dialed calls, thus the network resource can be utilized more effectively.

4.5.2 Cost-benefit analysis

iNetwork components are implemented as software components, and the connection of iNetwork components is IP-based. Constructing an iNetwork as an OGB telecommunication cost requires IP-based communication infrastructure. The most expensive component is BTS and the authority to use the frequency. Therefore, an organization must contract with GSM an operator to obtain the authority. However, the construction cost is an investment in the beginning of system operation. The telecommunication expenditure will be remarkable reduced thereafter.

Take an organization with 10 thousands members as an example. If every member needs to make 10 calls everyday, 30% of the calls are outgoing calls, others are connected to colleagues. Let 20% of the called colleagues are moving off the service region of the iNetwork community, other 80% can be reached in iNetwork service regions. Assume every call lasts for 2.5 minutes in average, and the calling fee per minute is 6 NT dollars. For every call, iNetwork can save the caller 8.4 NT dollars in average for every call.

70% between colleagues ´ 80% intra-community ´ 2.5 minutes per call ´ 6 NT dollars

From the view point of an organization, iNetwork can save the organization 840 thousand NT dollars every working day, and save 221,760,000 NT dollars for the organization every year.

8.4 dollars per call × 10 calls/per member per day × 10000 members = 840,000 NP dollars /day

From the above discussion, we found that the traffic load results from number translation queries are significantly reduced. The utilization of computing and transmission resource of the operator network will be enhanced. Implementing OGB network as a mobile PBX and applying OGB-based cache to mobile telecommunication systems is a cost-effective solution in alleviating mobile NP problem.

Chapter 5. Utilize 3G/WLAN dual-mode terminals to enhance the efficiency of mobile number portability service

5.1 Introduction

The results of chapter 3 and 4 exhibit that performing ported number translation in organization based (OGB) telecommunications network can substantially alleviate the workload and offered traffic of NPAC. Utilizing OGB cache to telecommunication NP service is a cost-effective approach for enhancing the efficiency of telecommunications NP service. However, OGB ported number translation benefits only the users in the service region of the organization network. Users exceed the service region still need to query NPAC for ported number translation. Maintaining routing information of ported numbers in OGB networks is not a comprehensive and satisfactory solution to NP services in FMC (fixed-mobile convergence) telecommunications environment.

A new opportunity for solving the NP problem presented along the emergence of 3G mobile telecommunications and the popular of intelligent mobile handsets. The property of data transmission channels and the ability of providing customized information allowing 3G mobile system to distribute personalized information to the user-end. The intelligent handsets with more storage and better computation capability enable the process of number translation in the user-end. The calling behavior of an individual user exhibits a strong locality of reference [5]. Caching the routing information of frequently dialed ported number in user-end can offload a substantial amount of traffic at the database. In addition, personalized information kept in a mobile handset (MH) carried by users can be utilized regardless of the physical location of the user. Because users often set up calls to the numbers stored in the address book of their mobile handsets, we propose to extend the address book of the dual-mode handset to keep the mapping of a user's frequently dialed numbers and the corresponding routing addresses, and to omit the time-consuming NPDB queries to mitigate the traffic load of NDAC.

To take advantage of user-end ported number translation to route the call directly without querying NPDB, the routing information of translated ported numbers shall be conveyed in the call-origination requests and be recognized by the public switching network. Also a mechanism which is convenient to distribute and renew the routing information to users' terminal is needed. The distribution of routing information of ported numbers shall be customizable and efficient without occupying computation and communication resources of telecommunication services. But providing personalized computing and customized data is costly in the 3G system. The communication model of 3G system is centralized. The knowledge for signal processing and information routing is concentrated in the core network [39]. The process and the transmission of customized information will consume considerable computational costs and network resources.

The development of communication systems is oriented toward heterogeneous network integration. Wireless LANs (WLAN) provide high throughput and broad capacity to users and considered a low cost solution for wireless access. The population of WLAN and 3G network systems brings the evolution of WLAN-3G integrated services, where WLAN is regarded as a complementary technology to 3G networks because of the limited coverage and considerable mobility management delay. WLAN-3G integration offers ubiquitous communications in high transmission rate, low communication cost, and also supports greater possibilities and opportunities for communication service providers. Customized applications and personalized services are feasible in WLAN-3G integration environment. The 3G/WLAN dual-mode mobile handsets (DMS) provide a way to combine the distributed data computational WLAN with knowledge centralized 3G networks in terminal and service level. In this chapter we describe the operation model of 3G/WLAN DMS with extended address books, and the method to communicate and transmit the confirmed routing information of ported numbers. We investigate the NP call setup time and the traffic load of NPAC to evaluate the performance of the purposed method.

On the other hand, a DMS contains the radio frequency (RF) modules of both WLAN and 3G systems that it can send and receive WLAN and 3G signals simultaneously. However, for the sake of small and intermittent WLAN service regions, DMS needs to handoff between WLAN and the 3G system frequently. Calling fee is a major consideration; hence, WLAN is in preference to the 3G system when choosing communication system. If a DMS originates a call via a congested WLAN, the fail of a try in WLAN brings extra effort to set up a call via 3G network and results in a long call-origination delay (COD). Consequently, DMS often confront the problems of long COD and high vertical handoff (VHO) overhead [43]. In addition, the increased signals and processes consume more battery energy. For eliminating the drawbacks of long COD, high VHO overhead, and high energy consumption, we propose an algorithm for DMS to determine which network system to engage to when originating a call or when entering a dual-mode communication environment.

The content of this chapter is organized as follows: In section 5.2 the mechanism to utilize address books of dual-mode MS to omitted NPDB queries is proposed. Also the cooperation and routing information distribution between the 3G network and WLAN are introduced. Section 5.3 is the DMS cost-efficiency evaluation. In section 5.4, the problems of using DMS in mobile telecommunications system are analyzed, and an algorithm to eliminating the drawbacks of DMS is proposed. Finally, the analysis model and the performance evaluation of the proposed algorithm is illustrated in section 5.5.

5.2 The implementation of dual-mode mobile NP service

While the dialed numbers of a user usually exhibit strong locality, we propose to enlarge the address book of users' MS to keep the routing information of a user's frequently dialed numbers. Thus, most of the ported numbers dialed by the user can be translated to the corresponding routing addresses in user's terminal without querying NPDB. The routing information must be borne in the call origination signals, and the signals must be recognizable to the public switching network; such that, the public switching network can route the calls to the appropriate networks directly without triggering the process of ported number translation.

For this purpose, the address book of MS shall be extended to keep the routing information, and the signaling system must be able to recognize the extra information in call-origination signals. Using user-provided routing information for call routing, the validity and the correctness of the information is important.

5.2.1 Utilize SS information to enhance mobile NP service

Every mobile network has a unique network ID that can be utilized to identify the subscription network of a portable or non-ported number. For keeping the routing information in the address book, every entry in the address book is extended to include the subscription network ID, as shown in Table 5-1.

The call origination procedure starts with a *SETUP* message sent from the user terminals to the switching network. The *SETUP* message carries the information of the called party's address and the supplementary service. The content of a *SETUP* message is listed in Table 5-2. When the address book of the dual-mode MS is extended to keep the subscription network of the mobile numbers, the *SETUP* messages shall carry corresponding information to the switching network. Where the *Facility* indicates that a Supplementary Service (SS) is enabled, the *SS version* is the version of the supplementary service—attaching the subscription network of the dialed number. The *Called party subaddress* can be used to indicate the subscription network of the dialed number.

Name	MSISDN	Subscription network ID
Alice	+8869315678	+88601
Benson	+8869264627	+88603
Carolyn	+8869192768	+88602
…	…	…

Table 5-1. Extra information in the address book of the dual-mode MS

The call is routed according to the subscription network ID. The switching network receives the SETUP message and finds that SS was enabled, that means the dialed address is translated. Then the call is routed to the subscription network of the dialed number directly. The subscription network of the dialed number consults HLR for the routing address of the termination network, and the call can be routed to the called party directly without querying NPDB.

This process can be implemented as a supplementary service of the 3G network, which omits routing messages to the NRH network and the time-consuming NPAC query; and it is backward-compliant with 2G and 2.5G systems. When more users dial the numbers in their address books, more workload of ported numbers translation is shifted to the user-terminal, and the efficiency of NP service is enhanced.

Information element	M: Mandatory O: Optional	Length (octet)
…		
Setup message type	M	1
Facility	**O**	**2-?**
Calling party subaddress	O	2-23
Called party subaddress	**O**	**2-23**
…		
SS version	**O**	**2-3**
…		

Table 5-2. The information elements of a *SETUP* message

5.2.2 Update cached data

The cached data must accurately represent the current subscription network of every frequently ported number in the DMS. A convenient and efficient mechanism to update the cached data is necessary.

Providing personalized computing and customized data is achievable but costly in the 3G system. The communication model of 3G system is centralized, and the knowledge for signal processing and information routing is concentrated in the core network [39]. The process and the transmission of customized information will consume considerable computational costs and network resources.

On the contrast, WLAN is designed as a packet-based data communication network which can also provide speech communications. It differs from telecommunication systems in the feature of distributed storage and computation. Users can access information and data of a WLAN by authentication process [40]. The distributed storage and computation properties of WLAN are appropriate to perform customized services in the client/server operation model [41]. The distribution and the update of routing information can be performed as an add-on service of WLAN that WLAN can distribute the routing information corresponding to the numbers in the address book of a user's MS.

A WLAN Routing Information Server (RIS) downloads the altered routing information from a neutral NPAC where a global NPDB is available, and RIS can distribute the routing information to the subscribers of the WLAN according to the previously registered user profile. Because the USIM (user service identity module) and the address book are shared resources of the 3G and WLAN systems [42], the pre-downloaded routing information in address books can be used to solve the routing address of ported numbers and mitigate NPDB queries in the two network systems. Based on USIM, WLAN can provide customized services to users.

Fig. 5-1 presents the operation of retrieving and distributing routing information in the 3G/WLAN dual-mode communication environment. All the routing information of every ported number is maintained in NPAC. When a number was ported out from the old subscription network, the old subscription network issued an update message to the NPAC for removing the related routing information from NPDB; when a number was ported to a new subscription network, the new subscription network requested the NPAC to update the new routing address of the number in NPDB. For the consistency of routing information to prevent loss of calls, usually the updated routing information is postponed for at least a couple hours to go into effect. Accordingly, the RIS of the WLAN updates the altered routing information periodically rather than instantly, and the process can be performed offline in the off-time.

Fig. 5-1 step 1 to 4 in white circles illustrates the process that a RIS retrieves altered routing information of ported number from NPAC. When a service provider network queries NPAC for altered routing information, NPAC determines the user authority and the time stamps to retrieve related information and sends it to the querying network.

Dual-mode MS users must interact with their subscription WLAN to update the routing information on their MS. A WLAN subscriber can register a profile of his address book in the RIS, thus the WLAN can provide customer-dependent services to the user. Fig. 5-1 step 1 to 6 in black circles illustrates the process of routing information update of a dual-mode MS. When a register message is originated from a dual-mode MS to the WLAN (step 1 in black circle), the message is forwarded to the Authentication Center (AUC) for user authentication (step 2). If the user is authorized, the AUC notifies the RIS to check whether the altered routing information is available for the user (step 3). The RIS filters the routing information according to the profile of the user (step 4), and issues a message to notify the MS of the user to update the altered routing information (step 5 and 6). Even when users move beyond the service region of their registered WLAN, they can access and update the renewed routing information if they can access the WLAN by the Internet or other WLAN.

The process of routing information update can be triggered when users turn on terminals in the service region of a WLAN, or when users connect their terminals to computers to synchronize data. The synchronization of routing information can be transferred through the Internet connection or via WLAN. The transfer speed of 802.11 a/g WLAN is 54 Mbps, which is more efficient than 384 kbps 3G radio access network (RAN).

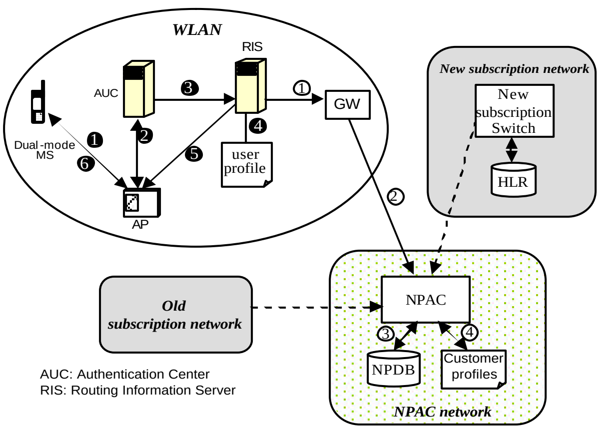

Fig. 5-1 The routing information retrieving and delivering for WLAN

5.2.3 The worse case

Even the update process can be carried out anytime when a user can access WLAN or Internet to prevent the routing information to be overdue. However, once networks are not available that the routing information on users' terminals cannot be updated, the obsolete routing information can lead to miss-routed calls. As shown in Fig. 5-2, the obsolete routing information will route a call to a wrong destination (step 1 to 4). The origination network needs to consult NPAC for the correct routing information of the dialed number (step 5 to 7), then uses the information to reach the subscription network of the called party to get the destination address to set up the call (step 8 to 11). This will result in the consumption of extra communication resources and the prolongation of the NP call setup time.

Though the call miss-routing is unpredictable and irresistible, the NP call service can not fail when the routing information on the user terminal is wrong. If a miss-routing happens, the calling network must be able to reissue a NPDB query to the callee's NRH network. For this purpose, the dialed phone number must be carried in the call origination message. Once if the routing information on user's terminal is obsolete, the origination network can query NPAC by the dialed number for obtaining the exact routing address of the callee's subscription network, then routes the call to the called party.

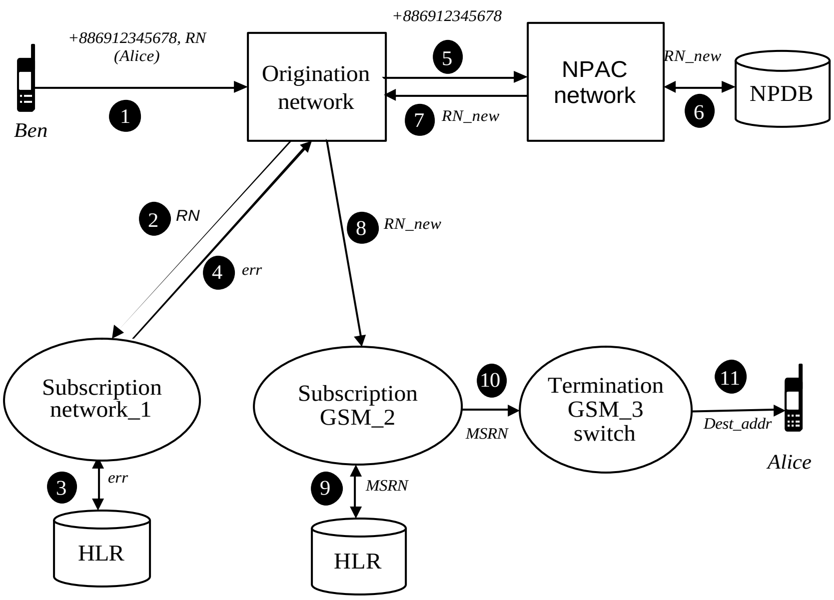

Fig. 5-2 The worst case of NP call routing

5.3 Performance and cost analysis of user-end caching

The routing information of ported numbers in the 3G/WLAN dual-mode MS allows most of the NP calls to be translated in user terminals, and the calls are routed to the subscription network directly without consulting NPDB. The response time of NP call setup is shorten as NPAC queries for ported number translation are omitted. While most of NPAC queries are omitted, the workload of NPAC is mitigated that can provide better performance of NP services to more customers.

We study the benefit of our method by evaluating the NP call setup time that is shortened, the NPAC traffic load be alleviated, and the average response time be saved for the NP call setup process.

5.3.1 The evaluation of call setup time and NPAC traffic load

While NPDB queries are all-call-based, every call set up to a number of the ported number-group initiates a NPDB query for number translation. The expansion of NP subscribers leads to an irresistible approach of huge NPDB and long searching delay. Here we evaluate the traffic load of NPAC and the NP call setup time to study the benefit of the proposed mechanism.

NP call setup time (denoted as t_c) includes the time for call processing (including signal processing, codec, channel reservation), call-setup signal transmission, and routing information query, that are respectively denoted as t_s, t_{trans}, and t_{query}. Where $t_c = t_s + t_{trans} + t_{query}$.

Consulting address books of dual-mode MS has the results of 1) cache miss, following the conventional NP call routing process; 2) cache hit and the data is correct, which can effectively reduce routing information query delay; and 3) cache hit but the data is overdue, which results in miss-routing and the worst case occurs. Let t_{NPAC} and t_{cache} represent the delay time of NPAC query and the address book query. Where $t_{NPAC} >> t_{cache}$, consequently the information-query delay in the conventional case is much longer than that of consulting address book ($t_{query\_conv} >> t_{cache}$).

- Case 1: When it is a cache miss, the NP call setup time (t_{c1}) of this case is very close to that of the conventional case because t_{cache} is very small that can be omitted. The time to query routing information is

  $t_{query\_1} = t_{cache} + t_{query\_conv}$,

  $t_{trans\_1} = t_{trans\_conv}$, and

  $t_{c1} = t_s + t_{trans\_1} + t_{query\_1}$.

- Case 2: Cache hits and the data is correct, the routing information query delay is

$t_{query\_2} = t_{cache} \ll t_{query\_1}$,

$t_{trans\_2} = t_{trans\_conv}$; the call setup time t_{c2} is

$t_{c2} = t_s + t_{trans\_2} + t_{query\_2}$, $t_{c2} \ll t_{c1}$.

- Case 3: Cache hits but the data is overdue, the call will be routed to a wrong address (the routing time is t_{trans_err}), and the call origination network has to re-issue a NPAC query for obtaining the exact routing information of the callee's subscription network. The time for call setup (t_{c3}) is

  $t_{c3} = t_{c2} + t_{trans\_err} + t_{trans\_conv} + t_{query\_conv}$.

  This is the worst case, t_{c3} is larger than ($t_{c2} + t_{c1}$). It is time consuming.

Reduce NPAC queries can alleviate both the call setup delay and the network resource consumption. Assuming the proportion of utilizing the routing information in address books to make NP calls is p, m proportion of the cached data is correct, and (1-m) will lead to miss-routing calls. The average call setup delay can be represented as

$(1 - p)\, t_{c1} + p\, [m \times t_{c2} + (1 - m)\, t_{c3}]$

Let 2% of the routing information in the address book is overdue. In a service area of the dual-mode communication environment with 10 thousands of users, every user generates 4 calls per day in average. Assume 30% of the calls are set to ported numbers, every routing information query is 1.5 seconds in average (including the query, transmission, and queuing delay) and 70% of the routing information is obtained form the address book. 7.2 hours information query delay is saved per day for the service area.

The NPAC traffic load is linearly proportion to the utilization of the address book of the dual-mode MS. Under the same condition, the extended address book can relieves the NPDB of about 14000 NPDB queries; that is, 35% NPDB traffic load is alleviated.

The more users obtain routing information from the address book of their MS, the more noticeable the NPAC workload is mitigated. Although parts of the local routing information is obsolete and causes miss-routing calls, utilizing the address book of dual-mode MS performs better NP service efficiency than the conventional NPAC queries do.

5.3.2 The evaluation of NP call setup delay

Considering NPAC as a single-server queuing system with a Poisson input of arrival rate λ and the service time is generally distributed with mean $t = \int_0^\infty t h(t) dt$, where h(t) is the density function of the service time. Assume the utilization factor r = ?t <1. We use M/G/1 model and the method of the embedded Markov chain to investigate the average queue length and waiting time. Let the queue length be N_q, the number of customers in the system is N, the waiting time is W, the average number of customers in the system is E[N] = E[N_q] + r. We have

$$E[N_q] = \frac{2r(1-r)+s}{2(1-r)} - r = \frac{s}{2(1-r)} ;$$

$$E[W] = E[N_q] / \lambda = \frac{s}{2l(1-r)} ,$$

where $s = l^2 \int_0^{\infty} t^2 h(t) dt$.

The relation of expected waiting time and the offered traffic of the NP service network is illustrated in Fig. 5-3. Let p denote the proportion of using routing information in MS instead of querying NPAC, and assume 2% of the local routing information is obsolete. When the offered traffic load is enormous that the limited service capacity of NPAC can not support the severe traffic load, the response time of routing information queries is prolonged remarkably that the performance of NP service becomes poor.

The routing information stored in users' MS alleviates the traffic of NPDB queries. With the same computation power, keeping routing information on users' MS can provide NP service to more users. Also the delay time of waiting on NPAC service can be reduced notably.

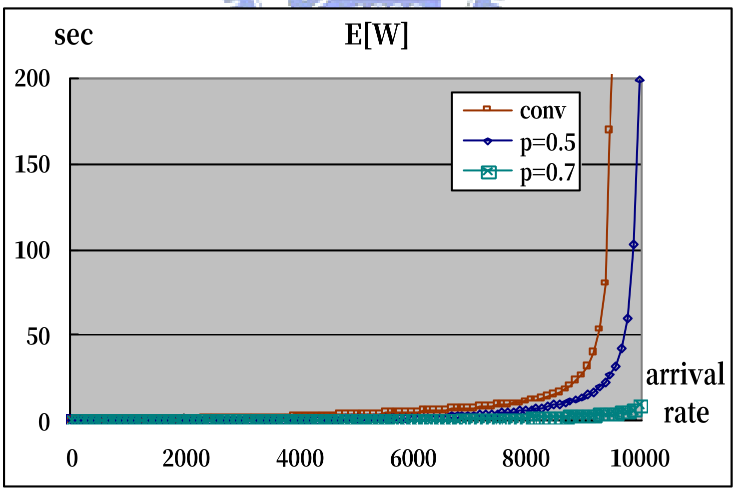

Fig. 5-3 The average waiting delay of a NP call setup process

5.4 Alleviate COD and VHO overhead of using 3G/WLAN DMS

The development of WLAN/3G DMS is an implementation of terminal- and service-level heterogeneous network convergence. With the ability to send and receive WLAN and 3G signals simultaneously, DMS combines the benefits of WLAN and ubiquitous 3G communication services in a single terminal. While 3G networks are expensive due to the high cost of spectrum acquisition, and the data rate is limited (up to 2Mbps), a DMS will first adopt WLAN for call origination in a dual-mode communication environment where a WLAN and a 3G network coexist to save communication cost. Besides, the energy for transmission in WLAN is half of that in a 3G network [43], adopt WLAN as the preferred network can extend the standby time of a DMS. When a user moves beyond the coverage of WLAN or when the visited WLAN is congested, the DMS adopts 3G system instead automatically. The determination and alternation of network adoption is transparent to users.

However, when a DMS originates a call via a congested WLAN, the fail of a try in WLAN brings extra effort to set up a call via 3G network and results in a long call-origination delay (COD). The traffic load of calls generated in a congested WLAN eventually offers to the 3G communication system. In addition, the increased signals and processes consume more battery energy. Dual-mode users who often originate calls via busy WLAN may spend a lot of time waiting on call setup and suffer from short standby time of DMS. To avoid the problems, 3G networks shall be given precedence over WLAN in congested hotspots.

For the sake of small and intermittent WLAN service regions, DMS needs to handoff between the 3G system and WLAN frequently. If a DMS moves out of WLAN coverage and enters a 3G cell, or vice versa, an ongoing connection shall be handed over between different network systems and causes a vertical handoff (VHO) process [44]. VHO is an expensive operation while it requires channel reallocations and information exchanges between heterogeneous network systems. If the dwell time of a DMS in a WLAN is short, the expense of VHO may nullify or exceed the communication profit from the connection in the WLAN. From the aspect of network systems, a dual-mode user who moves fast that the dwell time in a WLAN is too short to cover the cost of VHO shall keep on engaging to the visited 3G network rather than be handed over to the WLAN.

For eliminating the drawbacks of long COD, high VHO overhead, and high energy consumption, DMS shall be able to determine which network system to engage to when originating a call or when entering a dual-mode communication environment. But, without the circumstance information of how busy a WLAN is, how large the coverage of a hotspot is, how fast the user moves, and how long a connection will last, a DMS can not make determination. Network systems did not provide the circumstance information to users, therefore network systems and communication protocols should be modified to come out and to share the information with terminals. It is too expensive and too complicated to be a feasible approach.

Every dual-mode mobile user has his unique mobile and call patterns that can be induced from the history of the user's itineraries and the behavior of call originations [45]. The history data can be utilized to infer the condition of the visited communication circumstance, which is helpful in determining the preferred network to avoid the drawbacks of conventional dual-mode service.

5.4.1 Issues and problem analysis

The drawback of long COD and heavy VHO overhead happens when users enter or reside in dual-mode communication environments where WLAN and 3G networks coexist. Our approach is to avoid unnecessary attempts on WLAN call-generations in congested WLANs. Here we assume the coverage of WLAN hotspots are a subset of the service areas of 3G networks. The WLAN and the 3G modes are both actuated on a DMS.

The major factor which influences the success of call originations is the traffic intensity (or offered traffic) of a communication environment. A WLAN hotspot can be established in public and strategic areas by cellular operators for extending network coverage, or be established by organizations to facilitate internal communications or save communication costs. The traffic intensity of a WLAN varies with the amount of users, the location of the WLAN, and the observation time epoch. When the offered traffic load exceeds the carried load, new arrived calls are blocked or lost.

To prevent long COD, a mechanism to infer the communication condition of the visited WLAN before call setups is needed for a DMS. Currently user-ends cannot obtain information about the conditions of the visited communication circumstance. Neither WLAN nor the 3G network system provides a mechanism for sharing the information with user-ends. A DMS can only infer the communication condition of a visited circumstance from the history of a user's mobile and call patterns, including when a user used to visit a specific WLAN and how long he used to dwell in the WLAN; it can also calculate the frequency a user originated calls and the possibility of successful call-originations in a specific WLAN. By the gathered and calculated information, a DMS can determine whether it is proper to originate calls via the visited WLAN.

The possibility of successful call-originations decreases along the increase of the traffic intensity. Users experience the condition of the visited communication circumstance when originating calls. For representing the condition of the visited communication circumstance and how much a user relies on the WLAN, a DMS needs to keep the IDs of the visited WLANs in which calls were originated, the frequencies of call-originations in every visited WLAN, and the probability calls were successfully originated. The probability of successful call-originations in the peak time and the off time can be very different, the observation epoch must also be considered as a parameter of communication conditions.

From another point of view, the drawback of high VHO overhead results from the small and intermittent WLAN service regions. A connecting DMS moves across the border of 3G and WLAN networks causes VHO process to hand over an ongoing connection to a heterogeneous network. The WLAN communication profit from a user depends on how long the communication holds in the WLAN. If a hotspot is small, or a dual-mode user moves fast, the dwell time in the WLAN could be too short to cover the VHO cost. A connection shall not be handed over to WLAN when the VHO overhead is expected to be heavy. For users who have routine itineraries, DMS can learn the average call-holding time in a specific WLAN, and determine if VHO overhead is heavy. DMS can learn the average VHO overhead from users' moving patterns. A DMS notes VHO and the call lasting time in WLANs as information to denote the VHO overhead.

As a result, the parameters which represent the call and mobile patterns of dual-mode users, and the communication conditions of the visited WLAN can be concluded as the following:

1. The identifiers of visited WLANs (*SSID*). Every WLAN has a unique service set identifier (SSID). DMS keeps SSIDs of WLANs via which calls were originated.
2. The average call holding time in a WLAN (t_call).
3. The epoch to visit a WLAN (epochi). The grade of service (GoS) of a WLAN differs from the observation epoch. The time of a day can be divided into several epochs to distinguish the communication circumstances of every WLAN.
4. The frequency a DMS originated calls in the communication circumstance (calli).
5. The frequency of successful call-originations (scalli).
6. The last modified time of an entry (lmti) is aim for management purpose.

<table>
<tr><td rowspan="2">WALN ID ($SSID_i$)</td><td rowspan="2">Call lasting time (t_call_i)</td><td>$epoch_1$</td><td>$call_1$</td><td>$scall_1$</td><td>lmt_1</td></tr>
<tr><td>$epoch_2$</td><td>$call_2$</td><td>$scall_2$</td><td>lmt_2</td></tr>
</table>

Table 5-3 An entry of circumstance record on a DMS

We propose to maintain these parameters in users' DMS to log users' call and mobile habits, and to estimate the condition of the visited communication circumstance. The parameters relate to every communication circumstance are listed in Table 5-3. Since the memory of a DMS is limited, the oldest entry of the record will be removed when the memory runs out.

For avoiding long COD and frequent VHO, a set of rules to determine which network system to adopt in dual-mode environments is important.

5.4.2 Design of DMS training algorithm (DTA)

There are two reasons to give precedence of WLAN over 3G for dual-mode communication: The trials of call-originating in a congested WLAN must fail and must be regenerated via 3G network. That is, the traffic load generated by the calls happened in congested WLAN finally is offered to 3G networks, no matter WLAN or 3G is given precedence. Besides, the power consumption of DMS transmission in WLAN is half of that in 3G, WLAN is the preferred network of DMS in dual-mode communication environment.

Based on the concept of saving communication cost, a DMS adopts a WLAN for communication if the WLAN is first-time visited, and adds the circumstance information of the WLAN to DMS record. A modest communication circumstance represents high possibility of successful call originations, users are encouraged to utilize it for communication. In other respects, a user who often originates calls in a specific communication circumstance is considered as strongly relies on the circumstance. Even in the peak time, the user may like to originate calls via the WLAN to take chance. However, when users suffer from the frequent fails of the try, they are suggested adopting 3G network directly for shortening COD.

The coverage of WLAN is limited. While VHO is costly, frequent VHO is a notable drawback of dual-mode communication service. When the expected profit from a user will not exceed the operation cost of VHO, an ongoing call shall not be handed over to the visited WLAN. For a WLAN, the profit relates to the call holding time in the WLAN, which grows linearly as the total call lasting time. For users whose dwell time in WLAN is short shall adopt 3G network for communication. Thus the communication resource of WLAN and the 3G network can be utilized more efficiently.

The procedure of training a DMS to determine which network system to adopt is shown in Fig. 5-4.

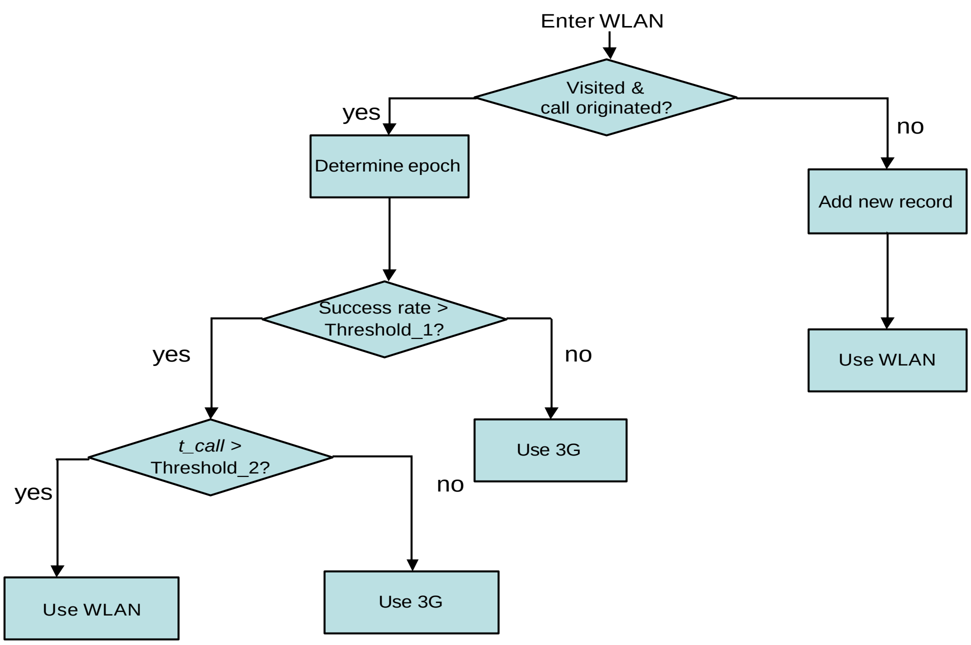

Fig. 5-4 The procedure for choosing a communication network

- Step 1: initialization

  A DMS checks the SSID of a WLAN after receiving the characteristics beacons of a WLAN and try to originate a call. The DMS increases the call-origination frequencies of the WLAN when a corresponding record exists; otherwise, it adds a new record for the WLAN.

```
If (ssid_i exists) {
    determine epoch j;
    call_j++;
} else {
   create a record of ssid_i;
   determine epoch j;
   call_j++;
}
```

- Step 2: determination:

  The rate of successful call originations represents the communication condition of a WLAN. In a modest WLAN, users are encouraged to utilize WLAN for communication. If a user often originated calls via the WLAN, it implies the user strongly relies on the WLAN. Even if the WLAN is busy, users may like to try the WLAN. In the two cases, WLAN is given precedence of call originations.

  From WLAN operators' viewpoint, if the expected revenue from a user may exceed the cost of VHO, the user is welcome to use the WLAN. For a user whose dwell time in a WLAN is too short to generate profit, the user shall keep on engaging to the 3G network rather than be handed over to the WLAN.

  The algorithm for determining which network system to adopt is depicted in the following:

```
If ((visit_i == 1) && (call_j >0)) {
    // The first time the user visit WLAN in epoch j, and the user has stayed for a long time
    Use WLAN to make a call;
    call_j ++;
} else if ((scall_j/call_j) > threshold_1) {
    // the rate of success-calls is high
    Use WLAN to make a call;
    call_j ++;
} else if (t_call_i > threshold_2) {
    // the call lasting time in this WLAN is long enough to cover VHO cost
    Use WLAN to make a call;
} else {
    Use 3G to make a call;
}
```

- Step 3: adaptation:

  After a call originated via WLAN is terminated, the DMS recalculate the average call holding time in the WLAN. Besides, an adaptation mechanism is invoked to update the record to present the condition of the communication circumstance.

```
If (success) {
   scall_j ++;
} else {
   // a chasten mechanism
   call_j --;
   scall_j --;
}
```

  While $call_j$ = $scall_j$, according to the inequality (y/x) = (y-1)/(x-1), where x,y>0 and x>y, decrease the value of ($scall_i$ /$call_i$) also decrease the opportunity of originating calls via WLAN.

For dual-mode users who have routine mobile and call patterns, the inferred possibility of successful call-originations and the call lasting time are close to the realities, thus the circumstance information will be really helpful. In some special cases such as parades or festivals, the traffic intensity increases suddenly that can not be inferred from the history data, and the DMS training algorithm (DTA) will not be helpful.

5.5 Analysis model and performance evaluation of DTA

5.5.1 Analysis model

DTA is applied to DMS to shorten COD in dual-mode communication environments, to decrease VHO overhead for communication systems, and to alleviate energy decline of DMS. The proposed DTA benefits the users who usually originate calls successfully or have long dwell time in WLAN. But it is unfavorable to the users who have short dwell time or who often visit WLAN in busy periods. By preventing call-originations that attempt to be failed in WLAN, the grade of service of WLAN is improved. Meanwhile, DTA avoids unnecessary VHO, thus the available channels can be utilized to higher profit communications.

We analyze the benefit of the proposed mechanism by evaluating the shortened COD and prevented unnecessary VHO.

■ Shortened COD in the user-end

Assume in the observation period of time t a WLAN is in the steady state. The average amount of users in the WLAN in (0, t) is m, and the call-origination rate of the i_{th} user is $?_i$.

When the traffic intensity exceeds the carried load of a WLAN, new arrived calls will be blocked or lost. Without the record in DMS and the training algorithm we proposed, the rate of successful call-originations in the WLAN is p_1, $0 \le p_1 \le 1$. Let the duration a user dwell in a WLAN is generally distributed. Let $g(x)$ denote the density function of the dwell time in a WLAN. The average dwell time of the i_{th} user is $t_{dwell} = \int_{x=0}^{t} x g(x) dx$. In this analysis, all the users are considered to generate the same pattern of traffic. The average traffic a user generates in a time unit is Erl. The carried load of the WLAN is $\sum_{i=0}^{m} p_1 \cdot l_i \cdot t_{dwell} \cdot Erl$. (5-1)

The calls failed to be originated via WLAN will be re-originated via a 3G network. The amount of re-originated calls (lost calls) is $\sum_{i=0}^{m} (1-p_1) \cdot l_i \cdot t_{dwell}$ (5-2)

By applying the proposed method to DMS, p_2 of the calls are originated via 3G network directly, where $0 \le p_2 \le 1$. The new traffic intensity of the WLAN becomes

$$\sum_{i=0}^{m} (1-p_2) \cdot l_i \cdot t_{dwell} \cdot Erl \qquad (5\text{-}3)$$

When the offered traffic load exceeds the carried load of the WLAN, the excess traffic load will be lost. The amount of lost calls can be represented as

$$\sum_{i=0}^{m} (1-p_1-p_2) \cdot l_i \cdot t_{dwell} \qquad (5\text{-}4)$$

Let the response time of call-origination in WLAN be t_{WLAN}, the total shortened COD = (reduced number of lost calls) $\times\, t_{WLAN}$ = $(\sum_{i=0}^{m} l_i \cdot p_2 \cdot t_{dwell}) \times t_{WLAN}$. (5-5)

■ Avoided VHO and broadened serviceable carried load in a WLAN

DTA can effectively avoid unnecessary VHO when dual-mode users have routine itineraries and call patterns. In addition, the transmission resources can be applied to high profit communications. In this chapter we investigate the amount of avoided VHO and the alleviated traffic intensity of WLAN to show the benefit of the proposed method.

Call arrivals and departures in a WLAN in the dwell time of an observer (t_{dwell}) are Poisson processes with call arrival rate *?* and departure rate *m*, respectively. When the offered traffic exceeds the capacity of a WLAN, the WLAN becomes congested, and new arrived calls will be blocked or lost. We use *r* to represent the utility of the WLAN, r = max (λ/μ, 1). In a WLAN, the offered traffic load is ?t_{dwell}, and the carried load is ρt_{dwell}.

According to DTA, a user whose dwell time in a WLAN which is too short will not be handed over from a 3G network to the WLAN. Assume the traffic load which can be alleviated by DTA is ?$_1 t_{dwell}$ in average, the offered traffic load becomes $(\lambda - \lambda_1)t_{dwell}$, and the number of lost or blocked calls can be represented as $(\lambda - \lambda_1 - \rho)t_{dwell}$.

Another important factor for measuring the performance of dual-mode service is the VHO overhead. As shown in Fig. 5-5 (a)-(c), a VHO (the shadows in Fig. 5-5) occurs when an ongoing call lasts to different network systems. Where ts_i and tt_i denote the time the i_{th} call began and terminated, to_i and th_i denote the time the observed DMS entered and left a WLAN, respectively. Calls arrive randomly in the observation period of time, thus the call holding time in a WLAN is proportional to the dwell time in a WLAN. Without applying DTA, VHO occurs when a call lasts to the service region of another communication system. The probability of VHO is p_{vho}= $P\{call\ lasting\ time > call\ holding\ time\ in\ a\ WLAN\} \cong P\{(tt_i - ts_i) > (th_i - to_i)/2\} = P\{tc_i > (th_i - to_i)/2\}$.

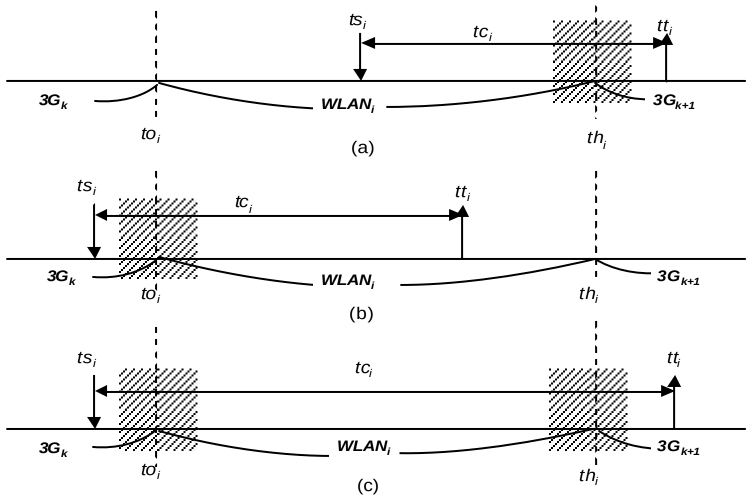

Fig. 5-5 VHO timing diagram

By applying DTA, when the average call holding time in a WLAN is inferred to be less than a threshold, the call will not be handed over to a WLAN. The amount of avoided VHO is $\lambda_1 t_{threshold}$. As a consequence, the potential traffic offered by the users is left out from the WLAN, and the extra transmission resource can be applied to other connections.

5.5.2 Performance evaluation

DTA allows DMS to infer the condition of a communication circumstance by the maintained history data, and to adopt the proper communication systems to shorten COD and to avoid unnecessary VHO. Thus, the WLAN communication resource can be utilized more effectively. For evaluating the performance of the proposed method, we estimate the amount of VHO and shortened COD when applying DTA, and compare the call loss rate of the purposed method and the conventional dual-mode communication service.

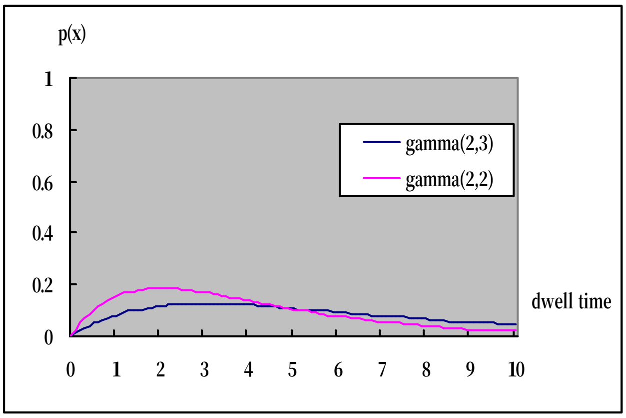

Fig. 5-6 Probability intensity function of users' dwell time

Mobile users may move through the coverage of a WLAN on foot or by vehicles. The resident time of a user in a WLAN ranges from seconds when moving by a high speed train to minutes for pedestrians, even lasts for hours in an office building. To represent the resident time of different kind of users in a WLAN, the distribution of a user's resident time is represented by a gamma distribution function. The probability intensity function is shown as in Fig. 5-6.

The proposed method of DTA sets a threshold of call holding time in a WLAN for preventing unnecessary VHO, where the value of threshold is set based on the average VHO delay or the accounting policy between WLAN and the 3G networks (e.g., 2 times of the VHO delay). When the average call holding time in a WLAN is too short to bear the VHO cost, the visited 3G network is adopted for call origination to prevent VHO. While a call in a WLAN is randomly started, the call holding time in a single WLAN is proportionally depends on the dwell time of a user in a WLAN.

The amount of traffic load which is generated by the avoided unnecessary VHO is shown in Fig. 5-7 and Fig. 5-8. Users whose average call holding time in a WLAN didn't exceed the threshold will not be handed over to the WLAN; thus, 10.5% and 28.2% of VHO can be avoided during off time and peak time, respectively.

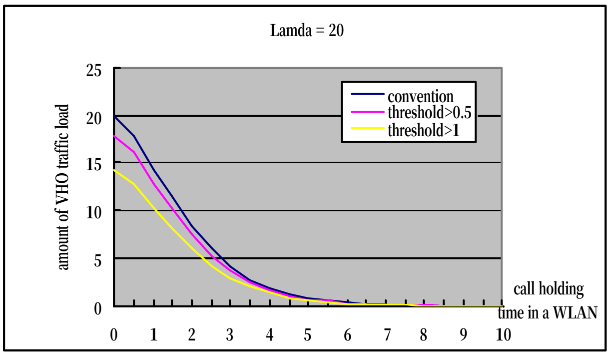

Fig. 5-7 The amount VHO traffic load (off time)

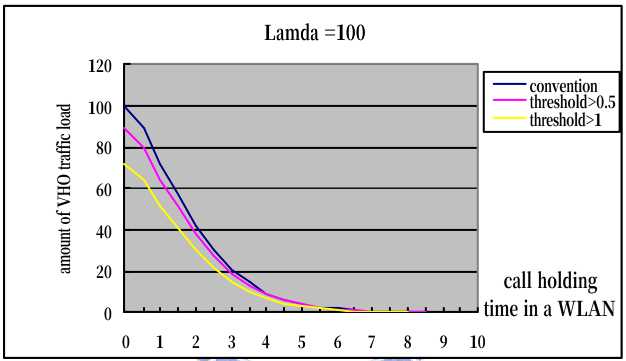

Fig. 5-8 The amount VHO traffic load (peak time)

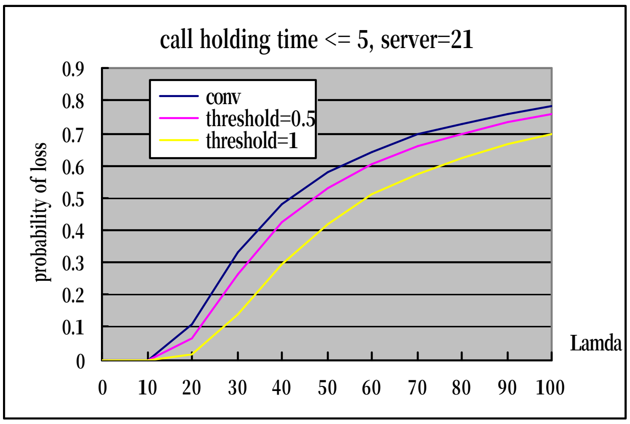

Fig. 5-9 The probability of loss

Users who adopt 3G network directly need not to wait for the fail of call-originations in WLAN, thus the COD can be shortened. Assume the response time of call-origination in WLAN (t_{WLAN}) is 1 second. The average amount of users in a hotspot of a train station is 30000 in peak time, the calling rate of a user in the

WLAN is 30% in average, and the rate of successful call origination is 70% without applying the proposed method. The success rate of call originations increases along the adaptation of DTA. When the 30% unsuccessful calls are originated from 3G network directly, the total shortened COD is 2700 seconds (45 minutes) during a single peak time period. For users who have routine itineraries and call patterns, the proposed method can save them remarkable waiting time and battery energy.

Users who often fail to originate calls via WLAN, or the call holding time in a WLAN is too short will be forced to adopt 3G network directly for communication by DTA. Thus, the congestion of busy WLAN can be alleviated and the call block and loss rate can be reduced. As shown in Fig. 5-9, DTA effectively reduces 10% of call losses in the peak time (100 call arrivals per sec), and 24.83% in the off time (50 call arrivals per sec).

5.6 Conclusion

Users usually have dial-number locality, they often utilize the address books in the MS. Keeping the routing information in the 3G/WLAN dual-mode MS (DMS) has the benefits of reducing the response time of routing information queries. The efficiency of telecommunications NP service is improved along the probability of utilizing the DMS address book and the cache hit rate increases. The routing information on users' terminal effectively alleviated traffic congestion and workload on NPAC to reduce the response time of ported number translation.

Our proposed method benefits both users and the NP service providers, which saves call setup time for users, and saves computing and communication resources for NP service providers. NP service providers can provide better performance to more users without extending of equipment and computing power. Without changing the existing signaling system and network architecture, the propose mechanism is a low-cost, effective, and efficient solution for improving the performance of mobile NP service.

On the other hand, long call-origination delay (COD) and high VHO overhead are the common drawbacks of WLAN/3G dual-mode mobile communication service. The drawbacks results in more battery energy consumption and shortened the standby time of dual-mode mobile handsets (DMS). In this chapter we proposed a service and terminal level solution to overcome the drawbacks. We study the mobile and calling behavior of dual-mode users to analyze the factors represented the condition of the communication environment and proposed to keep the factors in DMS. Also we propose a very simple algorithm, which is easy to implement on DMS, to infer the condition of the visited communication circumstance and to determine the preferred network to avoid the drawbacks of dual-mode services. Without changing existing communication architecture and protocols, the proposed method has the advantages of low complexity, easy to implement, and low-development cost.

We also evaluated the efficiency of the proposed method by establishing a simulation model to show that COD and VHO overhead of dual-mode communication service are alleviated effectively. Accordingly, the communication resource of WLAN and the 3G network can be utilized more efficiently, and the battery energy consumption is alleviated that the standby time of DMS is extended.

Chapter 6. Conclusion

6.1 Conclusion

Telecommunications technologies evolved from fixed-lined voice and data services to mobile multimedia communications. The long evolution of PSTN and the population of mobile communications system bring about the demand of fixed and mobile convergence (FMC) to supply a common platform for the intercommunications between fixed and mobile telecommunications users. In FMC environment, both the fixed-lined and mobile switching networks need to be able to determine and to route calls to the destination networks of the called parties.

Occurring with telecommunications liberation, more service providers join the telecommunications market, both for fixed and mobile telecommunications. Users have more choices and are more likely to change service providers according to the service, quality, and the billing policy offered by the operators. Number portability (NP) service allows a user to keep the same telephone number even when she changes operators. Calls set to a ported number can be routed to the subscription network and the destination address of the called party where the calling and called parties need not to sense about that. In the competitive telecommunications market, operators must provide NP service to attract subscribers and to enhance their competitiveness.

According to NP services, telecommunications systems need to maintain the mapping of subscriber numbers and the routing information to connect the subscriber. The large amount of NP users brings about the issue of long NPAC data searching delay that may bring about the accumulation of offered traffic in NPAC and result in long ported number translation delay to prolong the NP call setup time.

Based on the results of many researches, performing ported number translations in the early stage of NP call setup can effectively improve the efficiency of NP service. Caching the routing information of ported numbers for both fixed-lined and mobile telecommunications systems is an effective and feasible approach to solve the problems of serving NP functions in telecommunications environment. However, the size of cache and the update of cached data are important issues.

Cache performs well when the access of cached data has locality. By analyzing the hierarchy of telecommunications network, we found that the dialed numbers of users usually have locality. With respect to the evolution of the telecommunications environment, we proposed different cache-based solutions to alleviate the problem of telecommunications NP service. In the fixed-lined telecommunications system, such as PSTN, we proposed a PBX-based caching mechanism; in the mobile telecommunications systems before 2.5G, we proposed a mobile PBX telecommunication system and an approach of mobile PBX-based cache to alleviate the problem of NP service; in the mobile telecommunications system beyond 2.5G which possesses the capability of customized data service, we proposed to cache the routing information of an individual user's most frequently dialed ported numbers in the intelligent mobile handset.

According to the communication characteristics of users, we suggest adopting static cache policy. A static cache is established manually by system administrators in previous, rather than automatically performed by the communication system. The update of cached data is performed by IP-based networks. Hence, the update process can provide accurate routing information in time without occupying the telecommunication resource. The policy is simple and easy to implement, and the cache size is small enough to be arranged in memory that the turn-around time of cache consulting is very slight.

The most important considerations of caching mechanism are distributing the distribution of routing information of ported numbers, and allowing the public switching network to recognize the translated routing information. In this dissertation we proposed feasible and cost-effective mechanisms to fulfill the demands for different telecommunications environment. We also evaluated the efficiency and discussed the cost-benefit issues of the proposed method.

The improved NP call setup efficiency results in better communication resource serviceability. The call drop rate reduced remarkably along the decrease of call setup delay and the network resource can be effectively utilized. Therefore, our proposed methods benefit both users and the NP service providers, which can save call setup time for users, and save computing and communication resources for NP service providers. NP service providers can provide better performance to more users without extending of equipment and computing power. Without changing the existing signaling system and network architecture, the proposed mechanism is a low-cost, effective, and efficient solution for improving the performance of mobile NP service.

In addition, fixed-mobile convergence (FMC) is an evolution trend of data and telecommunications systems. User terminals which support FMC functions usually confront the issues of long call-origination delay (COD) and high VHO overhead. The issues will result in more battery energy consumption and shortened the standby time of user terminals. In this dissertation we proposed a service and terminal level solution to overcome the issues in the 3G-WLAN integrated communication environment. We study the mobile and calling behavior of 3G-WLAN dual-mode users to analyze the factors represented the condition of the communication environment and proposed to keep the factors in 3G/WLAN dual-mode mobile handsets (DMS). Also we propose a very simple algorithm, which is easy to implement on DMS, to infer the condition of the visited communication circumstance and to determine the preferred network to avoid the drawbacks of dual-mode services. Without changing existing communication architecture and protocols, the proposed method has the advantages of low complexity, easy to implement, and low-development cost.

The contributions of our works are:

- We proposed a novel approach to reduce the ported number translation delay in FMC telecommunications environment.
- With respect to the evolution of telecommunications systems, we proposed solutions to solve the NP problems in different telecommunications environment.
- The proposed solutions are easy to implement, cost-effective, and benefit users and service providers.

6.2 Future works

Along the population of NP service, users are more willing to engage to the services of integrated data and telecommunications systems, especially for the integrated IP-based and beyond 3G mobile communication systems. Many new applications and issues emerge from the development of services for such integrated communication systems. The following topics are the future works we are interested:

- Multimedia communications are the development trend of future telecommunications services. With respect to the transmission capacity and characters of different network systems, a mechanism to schedule and arrange the payload according to the property of the content to provide high quality communication service is necessary.
- Vertical handoff is inevitable in the integrated telecommunications environment. A comprehensive and effective scheme to mitigate the frequency of vertical handoffs and to reduce the handoff delay is urgent in the near future. The handoff delay consists of the latency of four stages of the handoff process:
    - The latency for scanning available systems and channels;
    - The reauthentication delay;
    - 4-way handshake delay (security key management); and
    - IP layer address translation.

  We will go into the research of smart handoff to reduce the handoff delay and handoff overhead.

References

[1] M. Foster, T. McGarry, J. Yu, *RFC 3482: Number Portability in the Global Switched Telephone Network (GSTN): An Overview*, 2003.

[2] Y. B. Lin, I. Chlamtac, H. C. Yu, *Mobile Number Portability*, IEEE Network, 2003.

[3] H. Y. Hsieh, K. H. KIM, and R. SIVAKUMAR, *An end-to-end approach for transparent mobility across heterogeneous wireless networks*, Mobile Networks ad Applications, pp. 363-378, 2004.

[4] J. H. Hu, G. Feng, K. L. Yeung, *Hierarchical Cache Design for Enhancing TCP over Heterogeneous Networks with Wired and Wireless Links*, IEEE Transactions on Wireless Communications, Vol. 2, No. 2, 2003.

[5] T. Caprenter, R. Carter, M. Cochinwala, M. Eiger, *Data Caching for Telephony Service*, International Database Engineering and Applications Symposium, pp. 101-109, 2002.

[6] M. Kihl, M. P. Rumsewicz, *On overload control of intelligent peripherals in intelligent networks*, Global Telecommunications Conference 1996, Nov. 1996, Vol. 3, pp. 1539 – 1543.

[7] J. Chen, J. Liao, and J. Chen, *Study on B-IP for provision of the multimedia video conference*, APCC/OECC '99. 5th Asia-Pacific Conference on Communications and Fourth Optoelectronics and Communications Conference, Oct. 1999, Vol. 2, pp. 884 – 887.

[8] S. McCann, H. Flygare, *Hiperlan/2 Public Access Interworking with 3G Cellular Systems*, Wireless Networks Vol. 10, No. 1, 2004, pp. 43-51.

[9] K. Lee, K. Kang, S. Kim, J. Lee, *Feature-oriented engineering of PBX software*, 6th Asia Pacific Software Engineering Conference, 1999, pp. 394-403.

[10] H. Wang, B. Raman et al., *ICEBERG: An Internet Core Network Architecture for Integrated Communications*, IEEE Personal Communications, Vol. 4, Aug. 2000, pp. 10-19.

[11] R. Jain, S. Rajagopalan, L. F. Chang, *A hashing scheme for phone number portability in PCS systems with ATM backbones*, 7th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1996, Vol. 2, pp. 593 –597.

[12] K. H. Kim, D. H. Lee, E. J. Ha, J. T. Park, J. H. Kim, S. K. Kim, *Provision of global number portability using CORBA*, IEEE/IFIP Network Operations and Management Symposium, 2000, pp. 17 – 30.

[13] M.A. Park, K.J. Park, S.K. Kim, Service management for IN-based local number portability, Fifth Asia-Pacific Conference and Fourth Optoelectronics and

Communications Conference on Communications, 1999, vol. 2, pp. 1290 – 1293.
[14] U. Black, The intelligent network, customizing telecommunication networks and services, Prentice Hall, 1998.
[15] Y. B. Lin, H. Rao, Number portability for telecommunication networks, IEEE Network, Vol. 13 Issue 1, Jan.-Feb. 1999, pp. 56 –62.
[16] H. S. Kim, H. S. Yong, "Personalized cache management for mobile computing environments", Information Processing Letters Vol. 87, Issue 4, Aug. 2003, pp. 221-228.
[17] http://www.arcome.com
[18] http://www.ida.gov.sg
[19] http://www.tct.hut.fi
[20] S. Khurana, A. Kahol, S. K. S. Gupta, P. K. Srimani, An efficient cache maintenance scheme for mobile environment, Proceedings of 20th International Conference on Distributed Computing Systems, 2000, pp. 530 – 537.
[21] B. Y. Chan, A. Si, H. V. Leong, Cache management for mobile databases: design and evaluation, Proceedings of IEEE CS Data Engineering, 1998, pp. 54-63.
[22] C. W. Pyo, J. Li, and H. Kameda, A caching scheme for dynamic location management in PCS networks, IEEE 58th Vehicular Technology Conference, 2003, Vol. 2 , pp. 761-765.
[23] J. Gans, S. King, G. Woodbridge, "*Numbers to the people: regulation, ownership and local number portability*", Information Economics and Policy Vol. 13, Issue 2, June, 2001, pp. 167-180.
[24] P. Faltstrom, "RFC 2916 – E.164 number and DNS", Sep. 2000.
[25] K. Lee, K. Kang, S. Kim, J. Lee. (1999). "Feature-oriented engineering of PBX software", Sixth Asia Pacific Software Engineering Conference, 1999, pp. 394-403.
[26] C. W. Cheng, C. G. Chung, "A PBX-based Approach for Telecommunication Number Portability Service", ICETE: International Conference on E-Business and telecommunication Networks, vol. 1, 2005, pp. 137-142.
[27] C. W. Cheng, C. G. Chung, "An Organization-based Cache Mechanism for Supporting PCS Number Portability Service", ICETE: International Conference on E-Business and telecommunication Networks, vol. 1, 2005, pp. 80-87.
[28] C. W. Cheng, C. G. Chung, "The design of a signaling system in an enterprise network system", International Conference on Communication Technology, 2003, Vol. 2 pp. 878-885.
[29] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: Session Initiation Protocol," RFC 2543, IETF, March 1999.
[30] C. W. Cheng, W. C. Liu, C. G. Chung, "iNetwork-A GSM compliant IP-based

Communication Platform", 25th Conference on Local Computer Networks, 2000, pp. 290 – 291.

[31] J. Kempf, R. Moats, P. St. Pierre, "Conversion of LDAP Schemas to and from SLP Templates", RFC 2926, Sep. 2000.

[32] E. Chlebus, W. Ludwin, "Is Handoff Traffic Really Poissonian?", Fourth IEEE International Conference on Universal Personal Communications, Nov. 1995, pp. 348 – 353.

[33] K.S. Trivedi, S. Dharmaaja, S. Ma, "Analytic modeling of handoffs in wireless cellular networks", Information Science Vol. 148, Dec. 2002, pp. 155-166.

[34] S. Dharmaraja, K.S. Trivedi, D. Logothetis, "Performance modeling of wireless networks with generally distributed handoff interarrival times", Computer Communications Vol. 26, Sep. 2003, pp. 1747-1755.

[35] U. Schoen, J. Hamann, A. Ugel, H. Kurzawa, C. Schmidt, "Convergence between public switching and the Internet", IEEE Communications Magazine, Vol. 36 Issue 1, Jan. 1998, pp. 50 –65.

[36] Y. B. Lin, I. Chlamtac, H. C. Yu, Mobile Number Portability, IEEE Network, 2003.

[37] H. Y. Hsieh, K. H. KIM, and R. SIVAKUMAR, An end-to-end approach for transparent mobility across heterogeneous wireless networks, Mobile Networks ad Applications, pp. 363-378, 2004.

[38] M.Cochinwala, Database Performance for Next Generation Telecommunications, 17^{th} International Conference on Data Engineering, 2001.

[39] M. Etoh, Next Generation Mobile Systems: 3G & Beyond, John Wiley & Sons, 2005.

[40] W.H. Yuen, R.D. Yates, and S. C. Mau, Noncooperative content distribution in mobile infostation networks, IEEE WCNC, Mar. 2003.

[41] D. Sánchez, E. Macías, A. Suárez, Effective Load Balancing on a LAN-WLAN Cluster, Parellel and Distributed Processing, Techniques and Applications, Vol. I, CSREA 2003, pp. 473-479.

[42] R. Rajavelsamy, V. Jeedigunta, B. Holur, M. Choudhary, and O. Song, Performance evaluation of VoIP over 3G-WLAN interworking system, Wireless Communications and Networking Conference, Volume 4, pp. 2312 – 2317, 2005.

[43] M. Nam, N. Choi, Y. Seok, and Y. Choi, "WISE: Energy-Efficiency Interface Selection on Vertical Handoff Between 3G Networks and WLANs", Personal, Indoor, and Mobile Radio Communications, PIMRC 2004, Volume 1, pp. 692 – 698, Sep. 2004.

[44] J. Korhonen, "Introduction to 3G mobile communications", Artech House, Inc., 2001.

[45] A. Klemm, C. Lindemann, M. Lohmann, "Traffic modeling and characterization for UMTS networks", GLOBECOM '01, vol 3, pp. 1741-1746, Nov. 2001.